Comparing Commuters' Short-Term and Long-Term Travel Mode Demand: Evidence from the Canton of Zurich

Dissertation
for the Faculty of Economics, Business Administration and Information Technology of the University of Zurich

to achieve the title of
Doctor of Economics

presented by
Oliver Bachmann
from Entlebuch

approved at the request of

Prof. Dr. Rainer Winkelmann
Prof. Michael Wolf, Ph.D.

The Faculty of Economics, Business Administration and Information Technology of the University of Zurich herewith permits the publication of the aforementioned dissertation without expressing any opinion on its views.

Zurich, December 5, 2007

The Dean: Prof. Dr. H.P. Wehrli

Acknowledgements

In writing this book, I have been fortunate to have had help and inspiration from different people. Since the beginning of this challenge, I have worked as a research assistant with Rainer Winkelmann, my thesis advisor. I want to thank him for his encouraging, helpful and open minded support throughout this period. He has read through this book and contributed with very constructive suggestions and criticism.

Moreover, I am grateful to Michael Wolf for co-advising this thesis. My sincere thanks also go to colleagues of mine from the Statistics and Empirical Research Group: Phillippe Mahler, Stephan Boes and Alejandra Cataneo. At regular intervals they gave uncountable many useful comments. Acknowledgements also go to Josef Troxler and Doris Rapold, both from the Statistical Office of Zurich, for making the data available and answering a number of data inquiries.

Last but not least, I am indebted to my parents and my brother who always supported and encouraged me without any ifs and buts.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Idea of this Thesis

Econometric research usually investigates commuters' travel mode choices independently from other decisions. That is, the individual travel mode decision is the only endogenous variable that is considered in the models, whereas all other variables associated with it are viewed as exogenous. As a consequence, the empirical results of such studies are only valid for a short time. This approach attributed to McFadden (1974) and Domencich and McFadden (1975) has a long history in transportation literature and planning, reflected by the mass of previous studies. Due to several advantages, including conceptual simplicity and excellent data availability, it has become the standard to analyze travel behavior (Ben-Akiva and Lerman, 1985).

In the long term, however, it is very likely that the selection of the travel means is closely related to other decisions. The recognition is not new that especially the choice of residence is a crucial aspect in the commuter's long-term travel mode choice process. Alonso's work (1964) provided the theoretical foundations, demonstrating that people face a trade-off between the prices of commuting and housing when choosing transportation means and residence. From this it follows that people with different prior travel propensities select themselves into residential areas that support their propensities in order to reduce their travel time and travel cost. In particular, as Cervero and Duncan (2002) argued, commuters who prefer traveling by rail tend to reside for that very reason near the railway station, while motorists have an incentive to live away from the station since areas near the station are often densely populated. Then spatial reallocation is possible as a result of changes in travel conditions, and commuters' long-term response to policy strategies can diverge from the short-term one (Boarnet and Crane, 2001; Romani et al., 2003; Vega and Reynolds-Feighan, 2005).

This thesis is the realization of the concept to study models that describe the short and long-term travel mode demand by the commuters in order to research in what the short and the long-term results differ. Hence, this work provides new aspects and insights that could be of interest to politicians and planners who intend to predict and evaluate the impacts of new traffic policy strategies. The short-term travel mode choice will be dealt with in Chapter 6. In order to obtain empirical findings that have a long-term validity, in Chapter 7 we investigate travel models where the commuter's home location appears endogenously. What deviations can we expect? Since the home location is endogenous in the long term, while being exogenous in the short term, there are more long-term than short-term opportunities to respond to changes in traffic policy strategies, and demand should therefore be more elastic in the long term.

1.2 Statistical Modeling

The travel mode decision in this work is a typical example of a qualitative choice. There are no quantities available as alternatives, but nominal values such as train, bus or car. For the description of such decision situations, the so-called qualitative choice models have asserted themselves in the econometric literature. They illustrate the individual decision process by means of a probability function.

In such models, the decision-maker is typically assumed to select the one alternative from a choice set which yields greatest utility, subject to some constraints. Since there are predictor variables that are observable by the analyst such as attributes of the alternatives and personal characteristics, and others that are not, utility consists of a deterministic and a random portion (the so-called error term). While typically not varying in the deterministic portion, qualitative choice models are characterized by varying specifications of the random portion. In Part I, the commonly used approaches and their use for our purposes are discussed in detail.

In modeling, there is usually a tradeoff between flexibility of the random portions and ease of estimation. The more the error terms allow describing travel behavior realistically, the more complex model estimation is. In the literature, for example, the multinomial logit is the standard model due to its intuitive closed-form formula that simplifies estimation (see Chapter 3). However, it may happen that the unobserved components of the utility function are heteroscedastic and/or correlated, a fact that violates the restrictive assumptions of the multinomial logit. In such cases, one must guess that the estimated results are biased. A reliable solution is to apply more flexible integral-form models such as the mixed multinomial logit (see Chapter 4) that, however, require simulation assisted estimation procedures. In recent years, such methodically advanced models have gained in popularity, in particular due to substantial progress in computer capacity and speed.

Part I also attaches importance to some theoretical aspects that have been discussed intensively in the latest econometric literature. For example, we address identification and normalization issues associated with mixed multinomial logit models, which have not been understood so far (see Chapter 4). We show that the number of identified error terms is not identical in mixed multinomial logit and in multinomial probit models, a fact that contradicts the prevailing assumption and practice published in literature. This is therefore the first study that presents the results of fully identified error component multinomial logit travel mode models that are estimated without need to implement simplifying a priori structures.

1.3 Why Using Swiss Census Data?

Our investigations are based on data gathered in the context of the Swiss census of the year 2000. In Chapter 5, we will thoroughly describe the data record and summarize both samples used in the empirical investigations. Before that, we would like to discuss a few general characteristics and specialties in short.

The data record contains disaggregated data from people living in Switzerland by December 5, 2000. The term *disaggregated* implies that the single person such as the commuter is the basic decision-maker unit. Observing each person only once, the data structure is cross sectional. Moreover, the data available is observational rather than experimental. More precisely, the travel mode variable states each commuter's revealed transportation preference on the regular journey to work. Preferences are revealed when the data represents the respondent's preferences in real activities. The benefit of revealed preferences is that they are critical for obtaining realistic choice information. However, this data also exhibits some drawbacks, including the potential presence of high correlation between predicting variables in real markets (for example between travel time and travel cost), leading to high standard errors in the estimations. In our research, this is not a problem since the high number of observations used in the data sample allow for highly precise estimations.

In the census data, you find the individual traffic means choice of the commuters, including additional information such as travel time for the chosen traffic means, commuting frequency, etc. In addition, the database also contains valuable information about the commuters themselves and their environments. A substantial problem, however, is the lack of central explanatory variables. In chapter 5, we thoroughly investigate the methods of how to extricate the lacking information endogenously from the data. As far as the data quality is concerned, the variables are generally complete, i.e. there are only few gaps. Overall, we can say that the Swiss census data is a good choice for our purposes, even if a few problems such as lacking variables have caused the author much work.

1.4 Why Studying the Canton of Zurich?

With an excellently designed street net and public transport network, the Canton of Zurich is literally predestined to investigate the commuting behavior of the working population. On a cantonal area of 1,728 squared km, the accumulated street length is more than 7,000 km and the accumulated length of the public transport means nearly 3,600 km, including 26 railways and more than 300 bus lines. For this work, good access to traffic is very important since we only regarded persons who had access to all three traffic means train, bus and car.

When we talk about commuter behavior in the canton of Zurich, we should look at some stylized facts. As the economic center of Switzerland, the canton of Zurich has experienced a significant increase of daily commuter flows in recent decades. A study by Frick et al. (2004) pointed out that the number of commuters more than doubled between 1970 and 2000, whereas the number of economically active people increased by only 27 percent in the same period. As a result, approximately 60 percent of the working population was leaving their residence commune in order to go to work at the end of the millennium.

At the same time, there was a strong reliance on the automobile as the primary transportation means. Census data 2000 unveils that almost 60 percent of all commuters living in the canton of Zurich and aged above eighteen years were car drivers, which amounts to approximately 1 percentage point less than ten years earlier[1]. With a noticeable plus of 3 percentage points since 1990, public transportation means such as rail, bus, tram and other were used by 36 percent. The remainder walked, rode by bike or motorbike, or went by company bus (minus 2 percentage points since 1990).

Within the public transport means, the percentage has changed massively between 1990 and 2000: While the demand for train connections has increased to 30.6 percent (plus 5 percentage points), the demand for bus/tram decreased to 4.8 percent (minus 2 percentage points). Frick et al. argued that

[1] Not counting the category ”no statements”.

rail popularity has been driven by numerous costly investments aiming at improving rail infrastructure and services, while the development of new bus services has been neglected a little. Overall, we can conclude that the use of fast transportation means tended to increase at the expense of slower transportation means.

1.5 Social Costs and Swiss Traffic Policy

In recent decades, there have been serious concerns in Switzerland about social costs mainly attributable to growing commuter activities. Typically, traffic experts view car use as the origin of strong inefficiencies. Each motorist weighs only his personal costs and his/her decision to drive may lead to negative externalities such as traffic jams, declining air quality, pollution, and noise, which are not reflected by market prices.

For the Canton of Zurich, there are no estimations about the amount of external costs; however, such estimations exist for Switzerland. According to Eichenberger and Schelker (2004), car traffic causes annual external costs of 5 to 8 billion Swiss Francs, depending on the way of calculation. Public transport also causes costs. Eichenberger and Schelker determine the value of state subsidies alone to 7 billion Swiss Francs annually, whereas train traffic causes the main load, approximately another 1 billion for environmental costs.

We must be conscious of the fact that commuter traffic causes much of the mentioned external costs, mainly during rush hour in the morning, at lunchtime and in the early evening. Consequently, contemporary Swiss traffic policies aim at altering the mode choice behavior during rush hour periods. In concrete terms, the policy aims at redirecting traffic away from the road onto the rails. The strategies to achieve this goal mainly include monetary disincentives for using the car in form of taxes, or time and monetary incentives for using the rail or other public transport in form of expensive infrastructure investments or the mentioned subsidies.

This study does not want to argue in favor or against of one traffic means. Instead, it provides necessary information required for implementing efficient traffic solutions by policy-makers. From a policy standpoint, understanding demand behavior is crucial for an accurate prediction and evaluation of the effectiveness of policy strategies aiming at mitigating costs caused by personal mobility.

1.6 Organization of this Book

The rest of this book is organized as follows: In Part I we address the current state of practice in qualitative choice modeling. This includes the foundations of qualitative choice modeling (Chapter 2), closed-form models such as the multinomial logit (Chapter 3), and integral-form models such as the multinomial probit and the mixed multinomial logit (Chapter 4). In Part II we study empirically workers' commuting behavior in the Canton of Zurich. This includes detailed data description (Chapter 5) as well as providing the results of the short-term (Chapter 6) and the long-term (Chapter 7) models. Chapter 8 gives a summary of the empirical findings and concludes with providing directions for further research.

Part I

Econometric Foundations of Qualitative Choice Modeling

Chapter 2

Qualitative Choice Models, Estimation, and Hypothesis Testing

2.1 Introduction

Qualitative choice models examine the relationship between an unordered qualitative choice variable on the one hand and several predictor variables on the other hand. As suggested by the terms *unordered* and *qualitative*, the values of the endogenous variable are labels. A commuter's choice between the travel alternatives rail, bus, and car is a typical example of an unordered qualitative variable. The variable takes the label *rail*, *bus*, or *car* if the commuter chooses the rail, the bus, or the car, respectively. For estimation, nevertheless, it is necessary to assign each label a number.

2.2 Random Utility Maximization and Qualitative Choice Models

We start with developing an analytical framework from which qualitative choice models can be developed straightforwardly. Based on random utility maximization (RUM), the concept presented in this section associates classical microeconomic demand theory with statistical randomness. It was originally put forward by Thurstone (1927) and further developed by Luce (1959) and Marschak (1960).

2.2.1 Choice Set

First we determine some basic requirements on the choice set from which the decision-maker selects an alternative. The choice set itemizes all alternatives that are optional in a certain choice situation. Formally, suppose $i = 1, ..., n$ individuals, each of them facing the same choice set C with $j = 1, ..., J$ alternatives. If choice sets differ over decision-makers, i's set is simply labeled C_i.

In general, researchers are free to define choice sets as long as the following properties are fulfilled:

1. The alternatives are mutually exclusive.
2. The choice set is exhaustive.
3. The number of alternatives is finite.

These conditions ensure that C fits into the modeling framework. (3) is the defining condition of qualitative choice models, while (1) and (2) require that each decision-maker selects one, and only one, alternative. To illustrate practical consequences of the properties above, study the set of transportation means $C =$\{rail,bus,car\}. In this example, the conditions are satisfied if each traveler in the data selected either the rail or the bus or the car.

2.2.2 Random Utility

From the decision-maker's perspective utility functions provided by the alternatives are completely deterministic. From our (the analyst's) perspective, however, choice reveals only which option is optimal. We can observe some but not all of the variables that influence utility. As a consequence, perfect prediction of individual behavior is impossible, and decision-maker's utility functions can be regarded as random variables.

Analytically, random utility of an alternative can be decomposed into two additive components. The observed portion takes the from of an indirect utility function, and the unobserved portion is specified as random variable, also denoted error term. Then the random utility of decision-maker i related to alternative $j \in C$ can be written as $U(x_{ij}, z_i, \epsilon_{ij}) = V(x_{ij}, z_i) + \epsilon_{ij}$, or briefly

$$U_{ij} = V_{ij} + \epsilon_{ij} \tag{2.1}$$

The indirect utility function V_{ij} includes the vector x_{ij} of observable attributes of the alternative and the vector z_i of observable socioeconomic characteristics of the decision-maker. The error term

ϵ_{ij} contains all unobserved variables affecting choice. In a model describing commuter's travel mode choice, x_{ij} usually includes time and cost of the trip, z_i involves income, age, sex, marital status, and so on, and ϵ_{ij} captures unobserved factors such as convenience and comfort of the alternatives when they are missing in the data sample.

To push the analysis beyond this very general illustration of the individual choice process, now we make concrete assumptions on the functional form of indirect utility. Keeping the model simple, indirect utility is normally specified to be linear in unknown population parameters[2], as follows:

$$U_{ij} = \alpha_{0j} + x'_{ij}\beta + z'_i\alpha_j + \epsilon_{ij} \tag{2.2}$$

where β and α_j are vectors of fixed population parameters to be estimated. β denotes a parameter vector associated with attributes of the alternatives, representing the population preferences towards attributes of the alternatives. α_j is a vector of parameters related to characteristics of the decision-maker, aiming at controlling observed choice heterogeneity in the population. Standard random utility representation also includes an alternative-specific constant term, which takes the value α_{0j} for alternative $j \in C$, and zero for all other alternatives $k \neq j \in C$, and which measures the effect of the dummy variable "alternative j".

An important note must be made here. In (2.2) the vector β is specified generically, i.e. there are no taste variations across alternatives. To allow for alternative-specific parameters, we can write

$$U_{ij} = \alpha_{0j} + x'_{ij}\beta_j + z'_i\alpha_j + \epsilon_{ij} \tag{2.3}$$

where $x_{ij} = 0$ for all alternatives $k \neq j \in C$.

2.2.3 Random Utility Maximization and Probability Model

In qualitative choice modeling, we assume that people behave rational. From the analyst's perspective the decision-maker selects the one alternative which yields maximal random utility, subject to the budget constraints. Formally, RUM implies that the alternative $j \in C$ will be chosen over all other alternatives if $U_{ij} > U_{ik}$ for all $k \neq j \in C$. From the random nature of the RUM process follows that individual choice behavior can be expressed as probability model. That is, the probability that a decision-maker i decides in favor of alternative j is formulated as the probability that the random

[2]Linearity in parameters does not imply linearity in variables.

utility provided by alternative j is maximal:

$$P(y_{ij} = 1) = Prob\{U_{ij} > U_{ik}, \forall k \neq j\} \tag{2.4}$$

The possible outcomes for the ith observation are given by the vector $y_i = (y_{i1}, ..., y_{iJ})$, where $y_{ij} = 1$ indicates that alternative j has been chosen by individual i, and $y_{ij} = 0$ means that alternative j has not been chosen by i. Keep in mind that each i selects one and only one alternative, thus $\sum_{j=1}^{J} y_{ij} = 1$ for all i.

We transform the inequality

$$P(y_{ij} = 1) = Prob\{\epsilon_{ik} - \epsilon_{ij} < V_{ij} - V_{ik}, \forall k \neq j\} \tag{2.5}$$

and receive a cumulative distribution function measuring the probability that the error differences are less than the differences of indirect utilities. The shape of the probability model is subject to the assumptions on the distribution of the error term differences. Once randomness is determined by the use of a specific probability distribution, the model can be derived from the RUM problem.

Using the linear indirect utility specification, we can rewrite (2.5):

$$P(y_{ij} = 1) = Prob\{\epsilon_{ik} - \epsilon_{ij} < \alpha_{0j} - \alpha_{0k} + (x_{ij} - x_{ik})'\beta + z_i'(\alpha_j - \alpha_k), \forall k \neq j\} \tag{2.6}$$

2.2.4 Identification and Normalization

There is no unique vector of parameters solving the unrestricted RUM problem in (2.6). The problem we face now is to determine the identified parameters and to impose normalization restrictions, which do not affect the choice probabilities while ensuring a unique solution of the maximization problem.

The most important identification rule follows from the fact that only differences in utilities matter for individual choice behavior. The relevant implication is that parameters are only identified if they capture differences across utilities of the alternatives (Train, 2003).

Indirect Utility Parameters

There are three groups of parameters appearing in indirect utility. Associated with attributes of the alternatives x_{ij} that include variation across the alternatives, the parameter vector β can be estimated without difficulty.

The vector of socioeconomic characteristics z_i, by contrast, does not differ across alternatives. Consequently, only the vector of parameter differences $\alpha_j - \alpha_k$, for all $k \neq j$, is identified. Because infinitely many vectors of parameter differences result in the same choice probability, we must further normalize the level of utility in order to guarantee a unique solution. We do this by restricting the parameter vector of one alternative to an arbitrary value, usually zero. The same is true for the vector of differences $\alpha_{0j} - \alpha_{0k}$, for all $k \neq j$. In order to set the utility level, the constant of one alternative is normalized to zero.

Error Term Parameters

The procedure of identifying and normalizing the parameters of the error term vector $\epsilon_i = (\epsilon_{i1}, ..., \epsilon_{iJ})$ usually involves the following two steps (Walker et al., 2004):

1. Reduce the variance-covariance matrix of the model, henceforth denoted Σ, by a transformation to the space of utility differences.

2. Apply the rank condition which states that the actual number of identified terms is equal to the number of independent terms in the transformation variance-covariance matrix $\Delta\Sigma$ minus one term that is used to set the scale of utility.

2.3 Maximum Likelihood and Maximum Simulated Likelihood Estimation

Qualitative choice models are usually estimated by maximum likelihood (ML) or, if the probability model has an integral-form, by maximum simulated likelihood (MSL). Both combine a simple estimation procedure with desirable large-sample properties of the estimators. ML estimation has enjoyed great popularity for a long time. In recent years, simulation assisted estimation methods like MSL have come into vogue as increased computational speed has allowed researchers to estimate richer models of consumer behavior.

2.3.1 Maximum Likelihood Estimation

Belonging to the class of extremum estimation procedures, ML estimation is defined through the maximization of a likelihood objective function. It is appropriate if the population distribution (i.e.

the data generating process) is known and if a sample of n observations has been drawn randomly from the population. Its basic idea is to determine the parameter estimates that maximize the probability of obtaining the observed sample.

Using the notation as previously, suppose a random sample with $i = 1, ..., n$ individuals, each provided with an outcome vector $y_i = (y_{i1}, ..., y_{iJ})$. Then the probability distribution across alternative outcomes of observation i can be represented by the multinomial distribution $f(y_{i1}, ..., y_{iJ}) = \prod_{j=1}^{J} P(y_{ij} = 1)^{y_{ij}}$, and the joint probability of n IID observations is expressed by $f(y_{11}, ..., y_{nJ}) = \prod_{i=1}^{n} \prod_{j=1}^{J} P(y_{ij} = 1)^{y_{ij}}$,

For a convenient illustration of the maximum likelihood procedure, let θ denote the overall vector of parameters, which contains all *identified* parameters of the random utility functions. Then the relevant likelihood function for the sample is defined as $L(\theta) = \prod_{i=1}^{n} \prod_{j=1}^{J} P(y_{ij} = 1|\theta)^{y_{ij}}$. Taking logs provides the familiar log-likelihood function

$$LL(\theta) = \sum_{i=1}^{n} \sum_{j=1}^{J} y_{ij} \log P(y_{ij} = 1|\theta) \tag{2.7}$$

The ML estimator of θ is the value that makes the joint distribution of the n IID observations most likely. Formally, it is defined as the value that maximizes the log-likelihood function:

$$\hat{\theta} = \arg\max_{\theta} LL(\theta) \tag{2.8}$$

ML estimation is feasible if $\log P(y_{ij} = 1|\theta)$ is analytically tractable, and computer routines can be written to evaluate this expression subject to any possible parameter vector θ. Make a notice that maximizing $L(\theta)$ leads to the same $\hat{\theta}$, but the logarithmic objective function simplifies the maximization problem mathematically.

At the true θ, the score function $S(\theta)$, defined as the derivative of the log-likelihood with respect to θ, is zero:

$$\begin{aligned}
0 &= \sum_{i=1}^{n} \sum_{j=1}^{J} y_{ij} \frac{1}{P(y_{ij} = 1|\theta)} \frac{\partial P(y_{ij} = 1|\theta)}{\partial \theta} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{J} (y_{ij} - P(y_{ij} = 1|\theta) + P(y_{ij} = 1|\theta)) \frac{1}{P(y_{ij} = 1|\theta)} \frac{\partial P(y_{ij} = 1|\theta)}{\partial \theta} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{J} (y_{ij} - P(y_{ij} = 1|\theta)) \frac{1}{P(y_{ij} = 1|\theta)} \frac{\partial P(y_{ij} = 1|\theta)}{\partial \theta} + \sum_{i=1}^{n} \sum_{j=1}^{J} (y_{ij} \frac{\partial P(y_{ij} = 1|\theta)}{\partial \theta} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{J} (y_{ij} - P(y_{ij} = 1|\theta)) \frac{1}{P(y_{ij} = 1|\theta)} \frac{\partial P(y_{ij} = 1|\theta)}{\partial \theta}
\end{aligned}$$

The last step follows because $\sum_{j=1}^{J} P(y_{ij} = 1|\theta) = 1$ implies $\sum_{j=1}^{J} \frac{\partial P(y_{ij}=1|\theta)}{\theta} = 0$, and therefore $\sum_{i=1}^{n} \sum_{j=1}^{J} \frac{\partial P(y_{ij}=1|\theta)}{\theta} = 0$. The term $\frac{1}{P(y_{ij}=1|\theta)} \frac{\partial P(y_{ij}=1|\theta)}{\partial\theta}$ denotes an instrument which is exogenous to $(y_{ij} - P(y_{ij} = 1|\theta))$.

This formulation of the first order condition of the ML estimation procedure provides an interesting interpretation. The difference between observation i's actual choice, y_{ij}, and the probability of that choice, $P(y_{ij} = 1|\theta)$, can be viewed as modeling error, or residual. For a random sample, the ML estimate $\hat{\theta}$ is therefore that value of the parameter vector θ that makes the residuals uncorrelated with the instrument.

2.3.2 Maximum Likelihood Estimation Properties

To derive the ML estimation properties, recognize that the score function is the sum of individual scores, thus $S(\theta) = \sum_{i=1}^{n} S_i(\theta)$ with $S_i(\theta) = \sum_{j=1}^{J} y_{ij} \frac{\partial P(y_{ij}=1)}{\partial\theta}$ varying over the n observations in the population. At the true parameter vector, $S(\theta)$ is distributed with $\mathrm{E}(S(\theta)) = 0$ and $\mathrm{Var}(S(\theta)) = \frac{\partial S(\theta)}{\partial\theta} = -\mathrm{E}(H(\theta))$, where $-\mathrm{E}(H(\theta))$ is the information matrix defined as minus the expectation of the Hessian matrix.

According to the central limit theorem, the sum of randomly distributed individual scores is normally distributed. Then the limiting distribution of the score function is given by

$$\sqrt{n}S(\theta) \rightarrow_d N(0, -n\mathrm{E}(H(\theta)))$$

The ML estimator $\hat{\theta}$ can be associated with the true parameter by taking the first order Taylor's expansion of $S(\hat{\theta})$ around $S(\theta)$: $S(\hat{\theta}) = S(\theta) + \frac{\partial S(\theta)}{\theta}(\hat{\theta} - \theta) = 0$. By rearranging the equation and multiplying both sides by $\sqrt{n}$, we obtain $\sqrt{n}(\hat{\theta} - \theta) = \sqrt{n}S(\theta)(-\frac{\partial S(\theta)}{\theta})^{-1}$, from which we can derive the asymptotic distribution of the ML estimator

$$\hat{\theta} \overset{a}{\sim} N(\theta, -\mathrm{E}(H(\theta))^{-1}) \tag{2.9}$$

Asymptotically, ML estimators are efficient among the set of consistent estimators in the context of a full parametric specified model.

2.3.3 Maximum Simulated Likelihood Estimation

The principle of MSL estimation is equal to that of ML. But unlike ML, the log-likelihood related to MSL estimation involves simulated probabilities:

$$LL(\theta) = \sum_{i=1}^{n} \sum_{j=1}^{J} y_{ij} \log \check{P}(y_{ij} = 1|\theta) \tag{2.10}$$

where $\check{P}(y_{ij} = 1|\theta)$ is the simulated probability that observation i chooses alternative j.

MSL estimation is used if choice probabilities do not have a closed-form expression. To understand this point, we rewrite the probability model in (2.6) by means of a $(J-1)$-dimensional integral

$$P(y_{ij} = 1) = \int \mathbf{1}(\epsilon_{ik} - \epsilon_{ij} < \alpha_{0j} - \alpha_{0k} + (x_{ij} - x_{ik})'\beta + z_i'(\alpha_j - \alpha_k, \forall k \neq j) f(\epsilon_{i,-j}) d\epsilon_{i,-j} \tag{2.11}$$

In this formulation, $\mathbf{1}()$ is an indicator function taking the value one if the statement in the bracket is true, and zero otherwise, and the error vector associated with observation i is defined as $\epsilon_{i,-j} = (\epsilon_{i1} - \epsilon_{ij}, ..., \epsilon_{ij-1} - \epsilon_{ij}, \epsilon_{ij+1} - \epsilon_{ij}, ..., \epsilon_{iJ} - \epsilon_{ij})$. Assumptions about the density function $f(\epsilon_{i,-j})$ specify whether the integral can be solved analytically, resulting in a closed-form probability expression, or not. If not, the computation of a standard log-likelihood function is infeasible on grounds of the inability of statistical software to compute the integral, and simulation assistance is required.

2.3.4 Probability Simulation

A mass of econometric studies has addressed techniques for evaluating an integral-form probability, and there are a variety of simulators which have been used in the past to approximate single and high-dimensional integrals. Surveys of Hajivassiliou (1993), Hajivassiliou and Ruud (1994), Hajivassiliou, McFadden and Ruud (1996), and Stern (1997) provide overview and discussion of this branch of research.

The general idea behind simulation of an integral-form probability is that every integral over a density is a kind of average. Then expression (2.11) can be viewed as a continuous mean that can be approximated by a discrete mean over many randomly chosen points sampled from a population distribution.

Two simulators have been usually used in recent qualitative choice literature, including the frequency simulator (also denoted accept-reject simulator) suggested by Lerman and Manski (1981), McFadden (1989) and Borsch-Supan and Hajivassiliou (1993), and the GHK simulator proposed by Geweke (1989) and independently developed by Hajivassiliou (1990), and Keane (1994).

The frequency simulator was an early innovation in approximating choice probabilities. For each draw taken from density $f(\epsilon_{i,-j})$, the indicator in (2.11) takes the value one if the statement in the brackets is true, and zero otherwise. Calculating the indicator for many draws and averaging the results gives the simulated probability. However, frequency simulators were found to perform badly under some circumstances. Since not continuous in parameters, simulated probabilities can take the value zero or one for a finite number of draws if the actual probability is close to zero or one. Furthermore, it can be difficult to obtain convergence when simulated probabilities do not move from rejection to acception (or vice versa) in response to parameter changes.

One solution addressing these issues is increasing the number of draws during simulation. Another, less time-consuming approach proposed by McFadden (1989) is substituting the pure zero-one indicator by more smooth indicators, for example the logistical function. However, the drawback is that smooth indicators usually produce biased probability estimates.

GHK is the predominant simulator of multinomial probit choice probabilities, which avoids both issues of frequency simulation. It rests on the idea that the probability integral can be transformed into a closed-form portion that can be calculated numerically and a non-closed form portion that can be simulated easily. GHK simulation is well proven and tested. In various comparisons, it has confirmed its usefulness and relative accuracy and has been found in MC studies to outperform other MNP simulation procedures including accept-reject (Geweke et al., 1994; Hajivassiliou et al., 1996).

2.3.5 Monte-Carlo Draws

Probability simulation requires random sampling from density $f(\epsilon_{i,-j})$. The computer routines are defined as follows.

1. For each observation i, draw a set of $d = 1, ..., D$ uniform distributed J-dimensional random numbers $u_i^1, ..., u_i^D$.
2. Using the inverse of the cumulative distribution function of ϵ_i, F_ϵ^{-1}, the uniform draws $u_i^1, ..., u_i^D$ can be transformed into a set of random numbers $\epsilon_i^1, ..., \epsilon_i^D$ by computing $\epsilon_i^d = F_\epsilon^{-1}(u_i^d)$.
3. For each draw $d = 1, ..., D$, calculate $\epsilon_{i,-j}^d = (\epsilon_{i1}^d - \epsilon_{ij}^d, ..., \epsilon_{ij-1}^d - \epsilon_{ij}^d, \epsilon_{ij+1}^d - \epsilon_{ij}^d, ..., \epsilon_{iJ}^d - \epsilon_{ij}^d)$.

The process of computer based random drawing is called pseudo-random Monte-Carlo (PRMC) method. The term *pseudo-random* implies that the random points generated by the computer are not truly random. The independence of PRMC draws facilitates the derivation of the statistical properties of probability simulation. However, short PRMC sequences can display uneven coverage of the area of integration, leading to an undesirable large simulation variance. To avoid this problem, recent studies typically employed between 500 and 1000 repetitions to provide sufficient simulation precision (see, for example, Revelt and Train, 1998; Train, 1998).

Though computers are getting faster and faster, making use of large sequences comes at the cost of extended computation time. Latest improvements have therefore aimed at developing better sampling methods. One successful method to enhance the precision for a given number of draws is denoted quasi-random Monte-Carlo (QRMC) sampling, which induces negative correlation between successive Monte-Carlo numbers by selecting the points more systematically. The introduction of dependencies among draws improves the coverage of the area of integration. Thus, a smaller number of points is sufficient to get simulation precision than with using common independent PRMC sequences.

There are numerous procedures generating correlated QRMC draws, including antithetic draws suggested by Hammersley and Morton (1956), nets which are employed in McGrath (1970), (s,m,t)-nets suggested by Train and Sandor (2004), Halton draws developed by Halton (1960), and modified Latin hypercube sampling methods proposed by Hess, Polak and Train (2006). The QRMC sequence most widespread used is the Halton sequence, mainly because it has been found to greatly outperform PRMC numbers in the context of simulation based estimation, at least when the number of dimensions is small (Bhat, 2003; Train, 1999). For example, Train found the simulation variance to be lower with applying 100 Halton draws than with applying 1000 PRMC draws.

Halton draws are generated using a prime number $p > 1$ as their base. The idea is that the Halton sequence cycles every p elements, systematically filling in the empty spaces of the interval on the unit line. To obtain a sequence of points for a given population density, the inverse cumulative distribution is evaluated at each element of the sequence according to the directions above.

The one-dimensional Halton sequence is generated by choosing p, expanding the sequence of integers $0, 1, 2, 3, ...$ in terms of this base, creating base p decimal numbers, and finally converting these base p decimal numbers back to base 10. For example, consider the sequence $0, 1, 2, 3, 4, 5$ and the base 2. The integers are transformed into base 2 by the following calculations: $0 = 0 \cdot 2^0 = 0$, $1 = 1 \cdot 2^0 = 1$, $2 = 1 \cdot 2^1 + 0 \cdot 2^0 = 10$, $3 = 1 \cdot 2^1 + 1 \cdot 2^0 = 11$, $4 = 1 \cdot 2^2 + 0 \cdot 2^1 + 0 \cdot 2^0 = 100$, $5 = 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 = 101$. Therefore, the digitized form of the sequence in base 2 is $0, 1, 10, 11, 100, 101$. Next, base 2 decimal numbers are created by reflecting the numbers about the decimal point: $0, 0.1, 0.01, 0.11, 0.001, 0.101$. Finally, base 2 decimal numbers are converted back to base 10, $0 = 0 \cdot 2^{-1} = 0$, $0.1 = 1 \cdot 2^{-1} = \frac{1}{2}$, $0.01 = 0 \cdot 2^{-1} + 1 \cdot 2^{-2} = \frac{1}{4}$, $0.11 = 1 \cdot 2^{-1} + 1 \cdot 2^{-2} = \frac{3}{4}$, $0.001 = 0 \cdot 2^{-1} + 0 \cdot 2^{-2} + 1 \cdot 2^{-3} = \frac{1}{8}$,

$0.001 = 1 \cdot 2^{-1} + 0 \cdot 2^{-2} + 1 \cdot 2^{-3} = \frac{5}{8}$. The Halton sequence is finally given by the numbers $0, \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}$.

In order to extend Halton sequences to more than one dimension, we must assign one base to each dimension. However, it has been shown that there is a strong correlation between sequences if the simulation dimensions are high (Bhat, 2003). Then the use of other sample procedures such as scrambling the digits or Latin hypercube sampling methods helps avoiding undesirable correlation.

Another issue with using Halton draws is that they produce deterministic rather than random numbers. Without randomization Halton sequences do not permit the statistical determination of simulation bias and noise. Different methods to randomize Halton draws have been proposed to compare their performance to that of PMC sequences. For example, Tuffin (1996) developed a simple and popular method that can be described in the following way: Take a draw labeled u randomly from a standard uniform distribution, and add it to each element of the Halton sequence. If the resulting number exceeds one, subtract the value one. Otherwise leave the result unchanged. For example, is $u = 0.7$ and the Halton number 0.2, the randomized number becomes 0.9. Is $u = 0.7$ and the Halton number 0.6, the randomized number becomes $0.6 + 0.7 - 1 = 0.3$. Note that this shift by the amount u preserves the property of the sequence.

2.3.6 Properties of Maximum Simulated Likelihood Estimation

The simulated score function $\check{S}(\theta)$, which must be zero at the true θ, can be decomposed into three separate terms $\check{S}(\theta) = S(\theta) + \{\mathrm{E}(\check{S}(\theta) - S(\theta))\} + \{\check{S}(\theta) - \mathrm{E}(\check{S}(\theta)\}$, where $S(\theta)$ is the score function defined in the ML case, and $\{\mathrm{E}(\check{S}(\theta) - S(\theta))\}$ and $\{\check{S}(\theta) - \mathrm{E}(\check{S}(\theta)\}$ denote the bias and the noise arising from simulation. The simulation bias is represented by the expected difference between the simulated score function and the score function at the true parameter. And the simulation noise is expressed as the deviation of the simulated scores from their mean.

This motivates a convergence analysis in the sense of observing the behavior of the MSL estimators with respect to the number of repetitions of the simulation process. Considering the limiting distribution of the score function as much as the limiting distributions of the bias and the noise, Hajivassiliou and Ruud (1994) demonstrated that for a sufficiently many draws MSL estimators behave asymptotically like their ML counterparts. Stated more detailed, they established that the asymptotic properties of MSL estimation is affected by the relationship between the number of observations n and the number of draws D, as follows:

1. For a fixed number of draws D, MSL estimation is inconsistent, even if the number of obser-

vations goes to infinity.

2. If D rises slower than $\sqrt{n}$, MSL estimation is consistent, but not asymptotically normal.

3. If D rises faster than $\sqrt{n}$, then MSL is equivalent to ML.

2.4 Hypothesis Testing

One benefit of ML estimation and, if the number of draws made during simulation is large enough, MSL estimation is that their asymptotic properties provide the capacity to examine single and multiple parameter restrictions. Asymptotic normality of the estimators leads to convenient asymptotic distributions of the test statistics, and efficiency of the estimators ensures maximum significance.

2.4.1 Asymptotic t-Test (Wald-Test)

The asymptotic t-test is used to examine whether single linear parameter restrictions hold or not. Examples include hypotheses that a single element θ_l of the vector θ is equal to a hypothesized constant, or the sum of a number of θ_l's equals a hypothesized constant. Formally, the null hypothesis $H_0 : r\theta = c$ is tested against the two-sided alternative hypothesis $H_1 : r\theta \neq c$, where r is a $1 \times L$ vector of fixed constants representing the coefficients that define the linear combination of the entries in θ that are of interest, θ is the $L \times 1$ parameter vector, and c is the hypothesized scalar.

For example, if the first element in r is one and all other elements zero, the null hypothesis $H_0 : \theta_1 = c$ is tested against the alternative hypotheses $H_1 : \theta_1 \neq c$. If the first two elements in r take the values 2 and 5 and all other elements are zero, the parameters θ_1 and θ_2 of θ are tested to follow the restriction $H_0 : 2\theta_1 + 5\theta_2 = c$ against $H_1 : 2\theta_1 + 5\theta_2 \neq c$. Otherwise, the single restriction can also be tested one-sided, and formulation of the alternative hypothesis determines whether the rejection area is on the right side (if $H_1 : r\theta > c$) or on the left side (if $H_1 : r\theta < c$).

Based on the asymptotic properties of ML and MSL estimation, the t-test statistic is under H_0 asymptotically standard normally distributed

$$\frac{r\hat{\theta} - c}{\sqrt{\widehat{var(r\hat{\theta})}}} \overset{a}{\sim} N(0,1) \tag{2.12}$$

where $\widehat{var(r\hat{\theta})}$ is the estimated variance of $r\hat{\theta}$. However, it should be noted that inference from t-tests is only valid if the examined parameters are uncorrelated to each other.

2.4.2 Likelihood Ratio Test

Able to capitalize information about correlations among parameters, likelihood ratio tests are used to inspect single as much as multiple linear parameter restrictions. Examples of such tests include hypotheses that an individual element θ_l of θ is equal to a hypothesized constant, a subvector of the elements of θ equals to some hypothesized constants, the whole vector θ equals to a vector of hypothesized constants, or the sums of a number of θ_l's equal some hypothesized constants. Formally, the null hypothesis $H_0 : R\theta = c$ is tested against the alternative hypothesis $H_1 : R\theta \neq c$, where θ is the $L \times 1$ vector of population parameters as before, R denotes a $Q \times L$ matrix of fixed constants representing the coefficients that define the Q linear combinations of the entries in θ that are of interest, and c is a $Q \times 1$ vector of hypothesized constants.

To derive the test statistic, take the second order Taylor's expansion of the log-likelihood of the unrestricted model around the log-likelihood of the restricted model: $LL(R\hat{\theta}) = LL(c) + S(c)(R\hat{\theta} - c) + \frac{1}{2}\frac{\partial S(c)}{\partial c}(R\hat{\theta} - c)^2$, with $R\hat{\theta}$ signifying the vector of estimators in the unrestricted model. If the null hypothesis is true, $S(c) = 0$ and $2(LL(c) - LL(R\hat{\theta})) = -\frac{\partial S(c)}{\partial c}(R\hat{\theta} - c)^2$, where $\frac{\partial S(c)}{\partial c} = -\mathrm{E}(H(c))$. Given asymptotic normality of ML and MSL estimation, the likelihood ratio test statistic is asymptotically χ^2-distributed with Q degrees of freedom (which denote the number of restrictions)

$$-2(LL(c)) - LL(R\hat{\theta})) \overset{a}{\sim} \chi^2\langle Q\rangle \tag{2.13}$$

2.4.3 Likelihood Ratio Index

The likelihood ratio index is an approach to measure the goodness of fit of qualitative choice models. By construction, the log-likelihood value at convergence of the estimated model is compared to the log-likelihood value at convergence of the so-called naive specification where the parameter vector is restricted to the null vector.

Mathematically,

$$\rho^2 = 1 - \frac{LL(\hat{\theta})}{LL(\hat{\theta} = 0)} \tag{2.14}$$

where $0 \leq \rho^2 \leq 1$. Note that the naive model has no explanatory power, thus $\rho^2 = 0$. If the model perfectly fits the data, $\rho^2 = 1$.

Unfortunately, the likelihood ratio index exhibits some limitations. It is a monotonic function of the sample size and the number of parameters in the specification. Consequently, it only provides valid comparisons of models with the same number of the same number of observations and parameters. Alternatively, one can compute an adjusted likelihood ratio index that penalizes rich models by correcting for the number of parameters:

$$\bar{\rho}^2 = 1 - \frac{LL(\hat{\theta}) - L}{LL(\hat{\theta} = 0)} \tag{2.15}$$

where L is the number of unknown parameters in the model.

Recognize that there is no direct relationship between ρ^2 and R^2 of a classical linear regression model. While R^2 measures the share of the variation of the dependent variable that is explained by the model, the likelihood ration index does not provide similar interpretation. Both numbers measure the model fit, but differently. According to simulations by Domencich and McFadden (1975), ρ^2 ranging from 0.2 to 0.4 indicates extremely good model fit, equivalent to R^2 ranging from 0.7 to 0.9.

2.5 Behavioral Output, Aggregation, and Forecasting

The principal aim of qualitative choice modeling is to understand demand behavior by examining how qualitative choices are affected by observable variables of the alternatives and the decision-makers. Due to the typically non-linear nature of qualitative choice models, however, the estimated indirect utility parameters do not measure the direct effect of the predictors on choice probabilities. Driven by the objective to reveal the economic significance of the estimation results, numerous statistics have been developed, including probability effects, marginal probability effects, point elasticities of the probabilities, and marginal willingness to pay for an attribute.

2.5.1 Probability Effect and Marginal Probability Effect

To examine the responsiveness of individual choice probabilities to small changes in the level of explanatory variables, we can compute probability effects and marginal probability effects, respectively. Formally, the effect of a *discrete* one unit change in the l-th attribute of the vector x_{ij} on i's probability of alternative j can be approximated by the probability effect

$$\Delta P(y_{ij} = 1) = P(y_{ij} = 1|x_{ijl} + \Delta x_{ijl}) - P(y_{ij} = 1|x_{ijl}) \tag{2.16}$$

For example, suppose that gasoline price rise increases the cost of car travel by 0.20 Swiss franc per trip. Then the value computed in (2.16) and divided by five predicts the change of commuter's car use probability.

Similarly, the probability effect of the n-th characteristic of the vector z_i tells us how the individual i's choice probability of alternative j changes if i's characteristic is $z_{in} + 1$ instead of z_{in}:

$$\Delta P(y_{ij} = 1) = P(y_{ij} = 1|z_{in} + 1) - P(y_{ij} = 1|z_{in}) \tag{2.17}$$

For example, the probability effect of the zero-one dummy variable *male* indicates how much the choice probability of an alternative j would change if the decision-maker's gender was male instead of female.

By contrast, the marginal probability effect measures the effect of an infinitesimal change in a continuous variable. With respect to attribute x_{ijl} it is formulated as

$$\frac{\partial P(y_{ij} = 1)}{\partial x_{ijl}} = \frac{\partial P(y_{ij} = 1)}{\partial V(x_{ij}, z_i)} \frac{\partial V(x_{ij}, z_i)}{\partial x_{ijl}} \tag{2.18}$$

Accordingly, relating to decision-maker's characteristic z_{in} it is

$$\frac{\partial P(y_{ij} = 1)}{\partial Z_{in}} = \frac{\partial P(y_{ij} = 1)}{\partial V(x_{ij}, z_i)} \frac{\partial V(x_{ij}, z_i)}{\partial z_{in}} \tag{2.19}$$

2.5.2 Point Elasticity of the Probability

Point elasticities of the probabilities express individual responsiveness to small changes in alternative-specific attributes by way of convenient percentage statements. The own point elasticity of the probability measures the percentage change of the probability of choosing an alternative with respect to a one-percentage change in an attribute of the same alternative. Formally, the own point elasticity of the probability of decision-maker i and alternative j with respect to the l-th attribute of the vector x_{ij} is obtained by the expression

$$E_{x_{ijl}}^{P(y_{ij}=1)} = \frac{\partial P(y_{ij} = 1)}{\partial x_{ijl}} \frac{x_{ijl}}{P(y_{ij} = 1)} \tag{2.20}$$

The cross point elasticity of the probability measures the percentage change of the probability of an alternative regarding a one-percentage change in an attribute of another alternative. Thus, i's cross point elasticity of the probability of an alternative j uses the l-th attribute of the vector x_{ik} of a competing alternatives $k \neq j$:

$$E_{x_{ikl}}^{P(y_{ij}=1)} = \frac{\partial P(y_{ij} = 1)}{\partial x_{ikl}} \frac{x_{ikl}}{P(y_{ij} = 1)} \tag{2.21}$$

2.5.3 Marginal Willingness to Pay for an Attribute

Knowledge of the shape of the individual demand curve can also be won by determining the decision-maker's willingness to pay for an attribute which measures the value that someone places on an attribute of an alternative by evaluating the marginal rate of substitution between the attribute and the cost of this alternative. Formally, suppose the attribute l of the alternative j labeled x_{ijl} and the cost of this alternative termed x_{ijc}. Then the willingness to pay for x_{ijl} in terms of x_{ijc} is simply written as

$$\frac{\partial V_{ij}/\partial x_{ijl}}{\partial V_{ij}/\partial x_{ilc}} \tag{2.22}$$

Note that this formula does not depend on the probability model. Only the specification of indirect utilities matters for computation.

In transportation literature, the value of travel time savings is an important indicator used to analyze cost and benefit of new transport systems. Traffic policy and planning are usually interested in knowing how much time is worth to travelers before financing new investments in infrastructure.

2.5.4 Aggregation and Forecasting

In qualitative choice models, individuals, households, firms and so forth represent the basic decision-maker unit. The models are specified and estimated at the disaggregated level. In order to receive forecasts of the behavior of the whole population, the statistics above must be subjected to aggregation.

The following two strategies are frequently used in practice:

1. *Representative individual approach*: Aggregated statistics are obtained by calculating individual statistics for the so-called representative individual that is characterized by average variable values.

2. *Sample enumeration approach*: Aggregated statistics are computed by averaging individual statistics over all observations.

The benefit of the representative approach is its conceptual clearness. Once the representative individual is identified, calculation of population statistics is straightforward. However, this strategy ignores the non-linear nature of the probability models. The behavior of an average individual is rarely a good approximation of the average behavior in the population.

In contrast to that, sample enumeration is a popular approach considering the whole sample. However, it is sensitive to outliers. In this case, a solution is to take the median instead of the mean.

Chapter 3

Models with Closed Form

3.1 Multinomial Logit Model

The multinomial logit (MNL) model is the most widely used model to describe the outcome of choice situations with three or more qualitative alternatives. Its popularity is mainly attributed to several advantages, including concreteness of the closed-form choice probability formulas and simplicity of estimation. However, it exhibits the undesirable independence of irrelevant alternatives (IIA) property that leads to biased forecasts when it is not supported by the data.

The MNL model can be derived using various approaches. In his seminal book, Luce (1959) derived the model by explicitly imposing the IIA restriction to the selection probabilities. Even though Marschak (1960) has showed that Luce's approach is consistent with random utility maximization, the abdication of behavioral assumptions is a profound deficiency. For this reason, we develop MNL choice probabilities from a behavioral RUM framework and a set of error term assumptions, following McFadden (1973) and Ben-Akiva and Lerman (1985).

3.1.1 Assumption on the Error Terms

Suppose the random utility $U_{ij} = V_{ij} + \epsilon_{ij}$, $i = 1, ..., n$, $j = 1, ..., J$. The central assumption of the MNL model is that the vector of error terms $\epsilon_i = (\epsilon_{i1}, ..., \epsilon_{iJ})$ is IID extreme value type 1 (EV1) with location parameters $\tau_j = 0$ for all $j = 1, ..., J$ and a scale parameter μ.

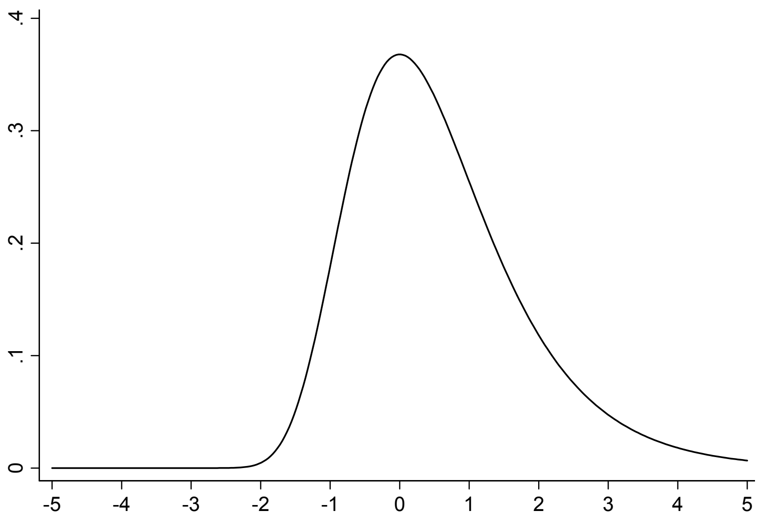
3.1.2 Extreme Value Distribution Type 1

A random variable ϵ_j has an EV1 (also denoted Weibull) distribution with a scale parameter μ and location parameter $\tau_j = 0$ if the probability density function is given by

$$f(\epsilon_j) = \mu e^{-\mu\epsilon_j} e^{-e^{-\mu\epsilon_j}} \tag{3.1}$$

Figure 3.1 plots the density function for $\mu = 1$, where $E(\epsilon_j) = 0.575$ and $Var(\epsilon_j) = \frac{\pi^2}{6}$. Skewed around the mean, the distribution displays thinner left and fatter right tails than a standard normal distribution.

Figure 3.1: Density function of EV1 variable



The gain of using IID EV1 variables is that the difference between two variables, $\epsilon_k - \epsilon_j$, is logistically distributed with cumulative distribution function

$$P(\epsilon_k - \epsilon_j \leq c) = \frac{1}{1 + e^{\mu c}} \tag{3.2}$$

where c is a constant.

Proof: Suppose $\epsilon_1 = \kappa +$
varphi and $\epsilon_2 = \kappa$. Then $\epsilon_1 - \epsilon_2 = \varphi$ for any κ and $f(\varphi) = \int_{-\infty}^{\infty} f(\kappa + \varphi) f(\kappa) d\kappa$. The antiderivative of the integral $\int_{-\infty}^{\infty} e^{\kappa+\varphi} e^{-e^{\kappa+\varphi}} e^{\kappa} e^{-e^{\kappa}} d\kappa$ is given by $e^{-e^{\kappa}(1+e^{\varphi})}(-\frac{e^{\kappa+\varphi}}{1+e^{\varphi}} - \frac{e^{\varphi}}{(1+e^{\varphi})^2})$. Thus, $f(\varphi) = -1 - \frac{e^{\varphi}}{(1+e^{\varphi})^2}$, which is the density function of the logistic distribution.

3.1.3 Choice Probabilities

When ϵ_{ij} is known, the probability model (2.5) can be written as

$$\begin{aligned} P(y_{ij}=1|\epsilon_{ij}) &= Prob\{\epsilon_{ik} < \epsilon_{ij} + V_{ij} - V_{ik}, \forall k \neq j\} \\ &= F_{\epsilon_{ik}}(\epsilon_{ij} + V_{ij} - V_{ik}, \forall k \neq j) \end{aligned}$$

which is the cumulative distribution function of ϵ_{ik} at $\epsilon_{ij} + V_{ij} - V_{ik}$ for all $k \neq j$.

Under the assumption of IID EV1 disturbances, the conditional probability becomes

$$\begin{aligned} P(y_{ij}=1|\epsilon_{ij}) &= \prod_{k\neq j} F_{\epsilon_{ik}}(\epsilon_{ij} + V_{ij} - V_{ik}) \\ &= \prod_{k\neq j} e^{-e^{-\mu(\epsilon_{ij}+V_{ij}-V_{ik})}} \end{aligned}$$

The unconditional choice probability is obtained by integrating $P(y_{ij}=1|\epsilon_{ij})$ over all possible values of ϵ_{ij}

$$\begin{aligned} P(y_{ij}=1) &= \int_{-\infty}^{+\infty} \prod_{k\neq j} F_{\epsilon_{ik}}(\epsilon_{ij} + V_{ij} - V_{ik}) f(\epsilon_{ij}) d\epsilon_{ij} \\ &= \int_{-\infty}^{+\infty} \prod_{k\neq j} e^{-e^{-\mu(\epsilon_{ij}+V_{ij}-V_{ik})}} e^{-e^{-\mu\epsilon_{ij}}} \mu e^{-\mu\epsilon_{ij}} d\epsilon_{ij} \\ &= \int_{-\infty}^{+\infty} \mu e^{-\mu\epsilon_{ij}} e^{-e^{-\mu\epsilon_{ij}}} e^{-\sum_{k\neq j} e^{-\mu(\epsilon_{ij}+V_{ij}-V_{ik})}} d\epsilon_{ij} \\ &= \int_{-\infty}^{+\infty} \mu e^{-\mu\epsilon_{ij}} e^{-e^{-\mu\epsilon_{ij}}} e^{-e^{-\mu\epsilon_{ij}} \sum_{k\neq j} e^{-\mu(V_{ij}-V_{ik})}} d\epsilon_{ij} \\ &= \int_{-\infty}^{+\infty} \mu e^{-\mu\epsilon_{ij}} e^{-e^{-\mu\epsilon_{ij}}(1+\sum_{k\neq j} e^{-\mu(V_{ij}-V_{ik})})} d\epsilon_{ij} \end{aligned}$$

For a finite μ and a constant c, $\int_{-\infty}^{+\infty} c\mu e^{-\mu\epsilon_{ij}} e^{-ce^{-\mu\epsilon_{ij}}} = 1$ and $\int_{-\infty}^{+\infty} \mu e^{-\mu\epsilon_{ij}} e^{-ce^{-\mu\epsilon_{ij}}} = 1/c$. Plugging $1 + \sum_{k\neq j} e^{-\mu(V_{ij}-V_{ik})}$ into c yields the choice probability in form of a logistic distribution

$$\begin{aligned} P(y_{ij}=1) &= \frac{1}{(1+\sum_{k\neq j} e^{-\mu(V_{ij}-V_{ik})})} \\ &= \frac{1}{(1+e^{-\mu V_{ij}} \sum_{k\neq j} e^{\mu V_{ik}})} \end{aligned}$$

Finally, the MNL selection probability of interest is

$$P(y_{ij}=1) = \frac{e^{\mu V_{ij}}}{\sum_{k=1}^{J} e^{\mu V_{ik}}} \tag{3.3}$$

The use of IID EV1 error terms leads to convenient closed-form probability formulas that are responsible for the long-standing popularity of the MNL model as benchmark model in the analysis of consumer's qualitative demand.

3.1.4 Identification of the Variance-Covariance Matrix

In Chapter 2.2.4, we have already addressed some general identification and normalization issues associated with random utility parameters. Now we discuss specific issues arising when estimating MNL models.

From assuming ϵ_{ij} IID EV1, $j = 1, ..., J$, it follows that the off-diagonal elements of the $J \times J$ variance-covariance matrix of the MNL model are zero and the diagonal elements are homoscedastic:

$$\Sigma = \frac{\pi^2}{6\mu^2} I$$

where I is the $J \times J$ identity matrix. Since the location parameter of the unobserved portion is zero by assumption, μ is the only parameter left to be determined.

For simplicity, suppose now a trinomial choice set, indexed by $j = 1, 2, 3$. Using the fact that only differences in utilities influence choice, we define $\Delta U_{ik,-1} = U_{ik} - U_{i1}$, $k = 2, 3$. The resultant 2×2 transformation variance-covariance matrix is given by

$$\Delta\Sigma = \begin{pmatrix} \frac{\pi^2}{3\mu^2} & \frac{\pi^2}{6\mu^2} \\ & \frac{\pi^2}{3\mu^2} \end{pmatrix}$$

$\Delta\Sigma$ has only one independent term that is used to set the scale of utility. According to the rank condition, there is therefore no room to estimate μ. In order to set the scale of utility, μ is usually normalized to one by multiplying the random utility functions by μ. Then the identified and normalized transformation matrix, henceforth denoted by means of an asterisk, is:

$$\Delta\Sigma^* = \begin{pmatrix} \frac{\pi^2}{3} & \frac{\pi^2}{6} \\ & \frac{\pi^2}{3} \end{pmatrix} \tag{3.4}$$

Be aware that the value of the scale parameter μ influences the MNL choice probabilities. If the scale parameter goes to zero, the deterministic part does not influence choice, and behavioral information is completely provided by the error term, since $\lim_{\mu \to 0} \mathrm{Var}(\epsilon_{ij}) = \infty$ and $\lim_{\mu \to 0} P(y_{ij} = 1) = \frac{1}{J}$. If the scale parameter goes to infinity, choice is completely deterministic, since $\lim_{\mu \to \infty} \mathrm{Var}(\epsilon_{ij}) = 0$ and $\lim_{\mu \to \infty} P(y_{ij} = 1) = 1$ if $V_{ij} > V_{ik}$, $\forall k \neq j$, and zero elsewise. Restricting μ to one is thus reasonable compromise between two limiting cases. Unfortunately, normalization comes at a cost. The indirect utility parameters are only identified together with μ. For example, only $\beta^* = \mu\beta$ can be estimated.

3.1.5 Maximum Likelihood Estimation

Now we describe the ML procedure to estimate the parameter vector θ of the MNL model. In general, ML easily estimates MNL models because closed-form choice probabilities permit a simulation-free computation of the score function and the Hessian matrix.

For a convenient illustration, we use previous notation. In the MNL, $\log P(y_i = j|\theta)$ is equal to $V_{ij} - \log \sum_k e^{V_{ik}}$, and the log-likelihood equation (2.7) can be reformulated as:

$$LL(\theta) = \sum_{i=1}^{I} \sum_{j=1}^{J} y_{ij}(V_{ij} - \log \sum_k e^{V_{ik}}) \tag{3.5}$$

The first order condition of the maximization problem is characterized by $S(\theta) = 0$. Thus,

$$\frac{\partial LL(\theta)}{\partial \theta} = \sum_i \sum_j (y_{ij} - P(y_{ij} = 1|\theta)) \frac{\partial V_{ij}}{\partial \theta} = 0 \tag{3.6}$$

It can be interpreted in the following way. The ML estimate is that value of the parameter vector θ that makes the residuals $(y_{ij} - P(y_{ij} = 1|\theta))$ uncorrelated with the instrument $\frac{\partial V_{ij}}{\partial \theta}$.

3.1.6 Probability Effect, Marginal Probability Effect, and Probability Elasticity

The MNL model is non-linear in indirect utility parameters. Hence, statistics such as probability effect, marginal probability effect and probability elasticity depend crucially on the utility level.

The probability effect is calculated if the change in the explanatory variable is discrete rather than continuous. Using formulas (2.16) and (2.17) in the context of a MNL, we obtain the probability effect of a one unit change in the l-th attribute of the vector x_{ij}

$$\Delta P(y_{ij} = 1) = \frac{e^{\alpha_{0j} + (x_{ij} + \Delta x_{ijl})'\beta + z_i'\alpha_j}}{\sum_{k=1}^{J} e^{\alpha_{0k} + (x_{ik} + \Delta x_{ijl})'\beta + z_i'\alpha_k}} - \frac{e^{\alpha_{0j} + x_{ij}'\beta + z_i'\alpha_j}}{\sum_{k=1}^{J} e^{\alpha_{0k} + x_{ik}'\beta + z_i'\alpha_k}} \tag{3.7}$$

and, respectively, the probability effect of a one unit change in the n-th characteristic of the vector z_i

$$\Delta P(y_{ij} = 1) = \frac{e^{\alpha_{0j} + x_{ij}'\beta + (z_i + \Delta z_{in})'\alpha_j}}{\sum_{k=1}^{J} e^{\alpha_{0k} + x_{ik}'\beta + (z_i + \Delta z_{in})'\alpha_k}} - \frac{e^{\alpha_{0j} + x_{ij}'\beta + z_i'\alpha_j}}{\sum_{k=1}^{J} e^{\alpha_{0k} + x_{ik}'\beta + z_i'\alpha_k}} \tag{3.8}$$

For continuous variables, the marginal probability effect is computed based on formulas (2.18) and (2.19):

$$\frac{\partial P(y_{ij}=1)}{\partial x_{ijl}} = \frac{\partial V_{ij}}{\partial x_{ijl}} P(y_{ij}=1)(1-P(y_{ij}=1)) \tag{3.9}$$

and, respectively,

$$\frac{\partial P(y_{ij}=1)}{\partial z_{in}} = P(y_{ij}=1)(\frac{\partial V_{ij}}{\partial z_{in}} - \sum_{k=2}^{J} P(y_{ik}=1)\frac{\partial V_{ik}}{\partial z_{in}}) \tag{3.10}$$

It is important to note that since both the probability effect and the marginal probability effect of a socioeconomic variable z_{il} are subject to the probabilities and parameters of all outcomes, their sign may differ from the parameter estimates (Winkelmann and Boes, 2005).

The formula of the own point elasticity of the probability of alternative j with respect to the continuous variables x_{ijl} is given by

$$E_{x_{ijl}}^{P(y_{ij}=1)} = \frac{\partial V_{ij}}{\partial x_{ijl}}(1-P(y_{ij}=1))x_{ij} \tag{3.11}$$

If an outcome gains (loses) market share[3], another outcome must lose (gain) market share ceteris paribus. As a consequence, the sign of the cross point probability elasticity of alternative j concerning the continuous variable x_{ikl} of a competing alternative k is negative:

$$E_{x_{ikl}}^{P(y_{ij}=1)} = -\frac{\partial V_{ij}}{\partial x_{ikl}} P(y_{ik}=1)x_{ik} \tag{3.12}$$

In the last expression, the probability and the attributes of alternative k appear, but not $P(y_{ij}=1)$. Thus, cross probability elasticities with respect to attribute x_{ikl} are identical when using the generic utility specification (2.2) with $\frac{\partial V_{ik}}{\partial x_{ikl}} = \beta_l$ for each outcome.

3.1.7 Limitations of Multinomial Logit Models

Due to IID error terms, the drawbacks with using the MNL model are that it does not allow for:

1. Unobserved heteroscedasticity and correlation between utilities
2. Unobserved taste heterogeneity in the population

[3]Probabilities can be interpreted as frequencies.

Both restrictions lead to one of the noteworthy aspects of the MNL model called the IIA property. This property stipulates that, for a given individual, the relative probability of choosing two existing alternatives (i.e. the odds-ratio) is unaffected by attributes or the presence of a third alternative. Formally, the odds ratio of two alternatives j and k does not contain elements of another alternative:

$$\frac{P_{ij}}{P_{ik}} = \frac{e^{V_{ij}}}{e^{V_{ik}}}$$

It is worth noting that the IIA property of the MNL model results from the IID and not from the EV1 distribution assumption. It is therefore present in any qualitative choice model with IID disturbances.

As a consequence of the IIA property, the MNL model predicts proportional substitution patterns between alternatives. This implies that a change in the attributes of one alternative changes the probability of the other alternatives proportionally. If some alternatives are closely related, MNL forecasts can be biased. Hence, the analyst should either perform an IIA test or employ less restricted models to examine whether the IIA property holds or not.

3.1.8 Hausman-McFadden Test of Model Structure

A widely used direct test of the IIA property has been proposed by Hausman and McFadden (1984). The idea of the test is as follows: As a result of the IIA property, the MNL probability odds for any two alternatives are the same whether or not other alternatives are available, and under the assumption of consistent estimation the parameter estimates obtained on a subset of alternatives $\tilde{C}$ do not significantly vary from those obtained on the full set of alternatives C. Thus, the advantage of this test is that it requires only estimating MNL models.

Formally, $H_0 : \theta_{\tilde{C}} = \theta_C$ is tested against the alternative hypothesis $H_1 : \theta_{\tilde{C}} \neq \theta_C$, where $\theta_{\tilde{C}}$ and θ_C are the vectors of parameters of the restricted and the unrestricted choice set, respectively. Under the null hypothesis, the Wald test statistic is asymptotically distributed χ^2 with Q degrees of freedom (which signify the number of parameters related to the excluded alternative)

$$(\hat{\theta}_{\tilde{C}} - \hat{\theta}_C)'(\mathrm{Var}(\hat{\theta}_{\tilde{C}}) - \mathrm{Var}(\hat{\theta}_C))^{-1}(\hat{\theta}_{\tilde{C}} - \hat{\theta}_C) \overset{a}{\sim} \chi^2\langle Q\rangle \tag{3.13}$$

For example, suppose the choice set C ={rail,bus,car}. If we suspect that unobserved correlation exists between the utilities of the alternatives train and bus, the Hausman-McFadden test contrasts the MNL estimates provided by C and the estimates provided by the restricted choice set $\tilde{C}$={train,car} (or $\tilde{C}$={bus,car}).

3.1.9 Binary Logit Model

Used to describe individual decisions from the binary outcome vector (y_{i1}, y_{i2}), the binary logit (BL) is a simplified version of the MNL. The probability formulas of observation i become

$$P(y_{i1} = 1) = Prob\{\epsilon_{i2} - \epsilon_{i1} < V_{i1} - V_{i2}\}$$

and $P(y_{i2} = 1) = 1 - P(y_{i1} = 1)$. From result (3.2) it follows that

$$\begin{aligned} P(y_{i1} = 1) &= \frac{1}{(1 + e^{-(V_{i1} - V_{i2})})} \\ &= \frac{e^{V_{i1}}}{(e^{V_{i1}} + e^{V_{i2}})} \end{aligned} \tag{3.14}$$

3.2 Generalized Extreme Value Models

Generalized extreme value (GEV) models relax the IIA property of the MNL model by allowing for unobserved correlation, while still maintaining the assumption that the error terms are identically distributed. Even though the number of possible models within the GEV class is theoretically limitless, only few forms prevail in today's qualitative choice literature. These include the nested MNL, the generalized nested MNL, and the cross-nested MNL model[4]. However, none of these models is flexible enough to approximate any RUM consistent behavior.

3.2.1 McFadden's Model Generating Process

GEV models can be generated straightforwardly from a non-negative function $G(w_1, ..., w_J)$ suggested by McFadden (1978). Following this methodology, G must satisfy three properties:

1. G must be homogeneous of degree $\mu > 0$, thus $G(aw) = a^{\mu} G(w)$
2. $\lim_{w_j \to \infty} G(w_1, ..., w_J) = \infty$ for each $j = 1, ..., J$.
3. The tth partial derivative with respect to t distinct w_j must be non-negative if t is odd and non-positive if t is even, that is for all distinct indices $j_1, ..., j_t \in \{1, ..., J\}$, we have $(-1)^t \frac{\partial^t G}{\partial w_{j_1} ... \partial w_{j_t}}(w) \leq 0 \; \forall w \in R^J_+$.

[4]Even though the MNL model belongs to the GEV family as special case, the term GEV is usually used to denote models that generalize the IID EV1 error terms.

The homogeneity condition attributed to Ben-Akiva and Francois (1983) is a generalization of McFadden's original unit homogeneity condition.

From G we can build:

- The probability model

$$P_j = \frac{w_j \frac{\partial G}{\partial w_j}(w_1, ..., w_J)}{\mu G(w_1, ..., w_J)} \tag{3.15}$$

- The expected maximum utility of the alternatives of a subset C_l

$$\mathrm{E}(\max_{j \in C_l} U_{jl}) = \frac{\log G(w_j, j \in C_l) + \lambda}{\mu_l} \tag{3.16}$$

  where λ is Euler's constant.

- The cumulative distribution function of the GEV error term vector

$$F(\epsilon_1, ..., \epsilon_J) = e^{-G(e^{-\epsilon_1}, ..., e^{-\epsilon_J})} \tag{3.17}$$

The use of McFadden's approach simplifies model derivation dramatically, yet provides only few insight in the decision-maker's choice process.

3.2.2 Nested Multinomial Logit Model

The most popular GEV model is the nested MNL (NMNL) model first derived by Williams (1977). The concept of this model is to divide the choice set into subgroups, denoted nests, and to allow the alternatives in the same nest sharing common unobserved factors among one another.

Figure 3.2 illustrates this idea graphically. Commuter's choice set C ={train,bus,car} is subdivided into nest C_1 ={train,bus} denoted public and nest C_2 ={car} denoted private. Including only one alternative, the private nest is called degenerated nest. In this example, rail and bus may be correlated, yet car and rail as well as car and bus may not.

The popularity of the NMNL model in recent decades has mainly been due to its intuitive conception and its closed-form probability formulas, which facilitate estimation. Nevertheless, the model's flexibility to describe unobserved correlation patterns is limited because each alternative may be member of only one nest.

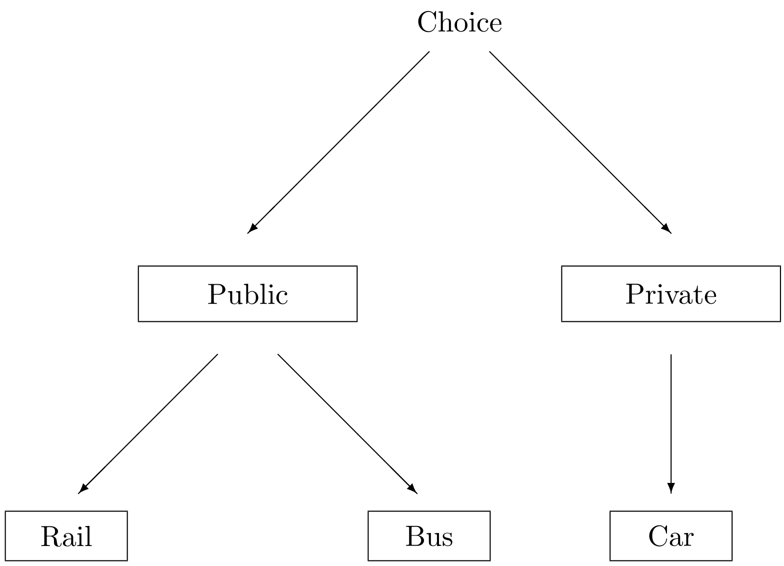


Figure 3.2: Nested structure

Choice Probabilities

Let each of the J alternatives of C be embedded in one and only one nest. The subset of alternatives associated with nest l is denoted C_l. Now suppose

$$G(w_1, ..., w_J) = \sum_{l=1}^{L} (\sum_{j \in C_l} w_j^{\mu_l})^{\frac{\mu}{\mu_l}}$$

If $\frac{\mu}{\mu_l} \leq 1$ for all $l = 1, ..., L$ and $G > 0$, G complies with theory:

1. $G(aw) = \sum_{l=1}^{L} (\sum_{j \in C_l} (aw_j)^{\mu_l})^{\frac{\mu}{\mu_l}} = a^{\mu} \sum_{l=1}^{L} \sum_{j \in C_l} w_j^{\mu_l})^{\frac{\mu}{\mu_l}}$.

2. $\lim_{w_j \to \infty} G(w) = \infty$, $w_j > 0$ $\forall j = 1, ..., \infty$.

3. $\frac{\partial G(w)}{\partial w_j} = \frac{\mu}{\mu_l} (\sum_{j \in C_l} w_j^{\mu_l})^{\frac{\mu}{\mu_l} - 1} \mu_l w_j^{\mu_l - 1} = \mu w_j^{\mu_l - 1} (\sum_{j \in C_l} w_j^{\mu_l})^{\frac{\mu}{\mu_l} - 1} \geq 0$, and
$\frac{\partial^2 G(w)}{\partial w_j \partial w_k} = (\frac{\mu}{\mu_l} - 1) \mu w_j^{\mu_l - 1} (\sum_{j \in C_l} w_j^{\mu_l})^{\frac{\mu}{\mu_l} - 2} \mu_l w_k^{\mu_l - 1} = \mu(\mu - \mu_l)(w_j w_k)^{\mu_l - 1} (\sum_{j \in C_l} w_j^{\mu_l})^{\frac{\mu}{\mu_l} - 2} \leq 0$.

From G we derive the NMNL probability model by applying the directions in (3.15)

$$P_j = \frac{w_j \mu w_j^{\mu_l - 1} (\sum_{j \in C_l} w_j^{\mu_l})^{\frac{\mu}{\mu_l} - 1}}{\mu \sum_{r=1}^{L} (\sum_{k \in C_r} w_k^{\mu_r})^{\frac{\mu}{\mu_r}}}$$

Plugging e^{V_j} into w_j, for all $j, ..., J$, gives the choice probability of alternative $j \in C_l$

$$\begin{aligned}
P_j &= \frac{e^{\mu_l V_j}(\sum_{j\in C_l} e^{\mu_l V_j})^{\frac{\mu}{\mu_l}-1}}{\sum_{r=1}^{L}(\sum_{k\in C_r} e^{\mu_r V_k})^{\frac{\mu}{\mu_r}}} \\
&= \frac{e^{\mu_l V_j}}{\sum_{j\in C_l} e^{\mu_l V_j}} \frac{(\sum_{j\in C_l} e^{\mu_l V_j})^{\frac{\mu}{\mu_l}}}{\sum_{r=1}^{L}(\sum_{k\in C_r} e^{\mu_r V_k})^{\frac{\mu}{\mu_r}}} \qquad (3.18)
\end{aligned}$$

In this formulation appearing, for example, in the popular textbook of Green (2003), the NMNL model is the product of two MNL models, including one describing the conditional probability of an alternative $j \in C_l$ given nest l, $P_{j|l}$, and one expressing the probability of picking up nest l, P_l.

Inclusive Value, Inclusive Value Parameter, and Identification

NMNL choice probabilities are sometimes stated in terms of an inclusive value first identified by Ben-Akiva (1973). The inclusive value of nest l is defined as

$$I_l = \log \sum_{j\in C_l} e^{\mu_l V_j} \qquad (3.19)$$

Since the expected maximum utility of nest l with subset C_l is

$$\mathrm{E}(\max_{j\in C_l} U_{jl}) = \frac{1}{\mu_l}(\log \sum_{j\in C_l} e^{\mu_l V_j} + \lambda)$$

the inclusive value of nest l can be interpreted as the summary of information about the alternatives within the nest (λ can be ignored on grounds of the irrelevance of the utility level for choice behavior).

With inclusive values, the NMNL choice probability of alternative $j \in C_l$ is reformulated as

$$P_j = \frac{e^{\mu_l V_j}}{e^{\mu_l I_l}} \frac{e^{\frac{\mu}{\mu_l} I_l}}{\sum_{r=1}^{L} e^{\frac{\mu}{\mu_r} I_r}} \qquad (3.20)$$

where $\frac{\mu}{\mu_l}$ denotes the inclusive value parameter of nest l. In this formula, the conditional MNL probability $P_{j|l}$ is proportional to the ratio of the utility of alternative $j \in C_l$ and the expected maximum utility of nest l. The marginal MNL probability P_l is proportional to the ratio of the expected maximum utility of nest l and the sum of expected maximum utilities over all nests.

As Ben-Akiva and Lerman (1985) established, correlation between the utilities of any pair of alternatives in the same nest l can be approximated by $1 - (\frac{\mu}{\mu_l})^2$. The degree of dependence between two alternatives within the same nest is summarized as follows:

- If $0 \leq \frac{\mu}{\mu_l} < 1$, there is positive correlation between the alternatives in nest l.
- If $\frac{\mu}{\mu_l} = 1$, there is no correlation between the alternatives in nest l.
- If $\frac{\mu}{\mu_l} > 1$, there is negative correlation between the alternatives in nest l.

The NMNL model is equal to the MNL if $\frac{\mu}{\mu_l} = 1$ for all $l = 1, ..., L$. As found by Train (2003), the NMNL model specification is consistent with decision-maker's RUM if $0 \leq \frac{\mu}{\mu_l} \leq 1$. For $\frac{\mu}{\mu_l} > 1$, the model is only consistent with RUM for some data ranges, while $\frac{\mu}{\mu_l} < 0$ is definitely inconsistent with RUM.

Identification and normalization of inclusive value parameters is straightforward. Since μ and μ_l are only identified together, it is common practice in the literature to normalize $\mu = 1$ in order to estimate the parameters $1/\mu_1, ..., 1/\mu_L$ from the data.

Simultaneous and Sequential Maximum Likelihood Estimation

Most statistical software packages contain ML routines for the simultaneous estimation of NMNL models. The term *simultaneous* means that the utility parameters of both decision levels of the decision tree in Figure 3.2 are estimated jointly. If simultaneous estimation is computationally burdensome, sequential ML estimation is a feasible solution. In this procedure, the estimates of the conditional MNL probability (lower model) are used to calculate the inclusive value parameter for each lower nest. Then the marginal probability (upper model) is estimated using the inclusive values as predictors.

The log-likelihood of sequential ML estimation is expressed as follows:

$$\begin{aligned} LL(\theta) &= \sum_{i=1}^{I} \sum_{j,l \in C} y_{ilj} \log(P_{ij|l}(\theta) P_{il}(\theta)) \\ &= \sum_{i=1}^{I} \sum_{j,l \in C} y_{ilj} \log P_{ij|l}(\theta) + \sum_{i=1}^{I} \sum_{j,l \in C} y_{ilj} \log P_{il}(\theta) \end{aligned} \tag{3.21}$$

If alternative $j \in C_l$ is chosen, $y_{ilj} = 1$, and $y_{ilj} = 0$ otherwise. Recognize that sequential estimation of MNL models can be inefficient. The loss of efficiency emerges when some of the unknown parameters appear on both estimation levels. Then sequential estimation does not utilize all information provided by the data.

Independence of Irrelevant Alternatives Property

In the NMNL, the IIA property holds within nests, but not in general across nests. Formally, assume two alternatives j and k, both arranged in the same nest l. Then

$$\frac{P_{ij}}{P_{ik}} = \frac{e^{V_{ij}/\frac{\mu}{\mu_l}}}{e^{V_{ik}/\frac{\mu}{\mu_l}}}$$

However, for any two alternatives j and k in different nests l and r, the odds ratio usually depends on the attributes or the existence of other alternatives in these two nests. In Figure 3.2, for example, the odds ratio between rail and bus is completely unaffected by the attributes or presence of car, while the odds ratio between rail and car can be affected by the attributes or the presence of bus. As a consequence, rail and car can display disproportional substitution patterns while rail and bus exhibit proportional substitution patterns by assumption.

Till the early 1990s, NMNL models have been viewed as a sound answer addressing the IIA. However, they provide only some relaxations from proportional substitution patterns.

3.2.3 Cross-Nested and Generalized Nested Multinomial Logit Models

The cross-nested MNL (CNMNL) model proposed by Ben-Akiva and Bierlaire (1999) and derived by Bierlaire (2001) is the most general member of the GEV model family thus far and has recently been given some notice in the qualitative choice literature. It includes as special cases the generalized nested MNL (GNMNL) model going back to Koppelman and Wen (2001), a preceding version of the CNMNL model suggested by Vovsha (1997), and the NMNL model discussed before.

The CNMNL model is characterized by allowing each alternative to be a member of more than one nest. Denote s_{jl} the share of alternative j in nest l, satisfying $\sum_{l=1}^{L} s_{lj} = 1$ for each alternative $j = 1, ..., J$. Then the CNMNL model is a GEV model derived from the function

$$G(w_1, ..., w_J) = \sum_{l=1}^{L} (\sum_{j \in C_l} s_{lj} w_j^{\mu_l})^{\frac{\mu}{\mu_l}}$$

The required properties are verified in the same way as previously. If $\frac{\mu}{\mu_l} \leq 1$ for all $l = 1, ..., L$ and $G > 0$, G complies with the theory because

1. $G(aw) = \sum_{l=1}^{L} (\sum_{j \in C_l} s_{lj} (aw_j)^{\mu_l})^{\frac{\mu}{\mu_l}} = a^{\mu} \sum_{l=1}^{L} (\sum_{j \in C_l} s_{lj} w_j^{\mu_l})^{\frac{\mu}{\mu_l}}$

2. $\lim_{w_j \to \infty} G(w) = \infty$, $w_j > 0 \forall j = 1, ..., \infty$

3. $\frac{\partial G(w)}{\partial w_j} = \sum_{l=1}^{L} \frac{\mu}{\mu_l} (\sum_{j \in C_l} s_{lj} w_j^{\mu_l})^{\frac{\mu}{\mu_l}-1} \mu_l s_{lj} w_j^{\mu_l - 1} = \mu \sum_{l=1}^{L} s_{lj} w_j^{\mu_l - 1} (\sum_{j \in C_l} s_{lj} w_j^{\mu_l})^{\frac{\mu}{\mu_l}-1} \geq 0,$
$\frac{\partial^2 G(w)}{\partial w_j \partial w_k} = \mu \sum_{l=1}^{L} (\frac{\mu}{\mu_l} - 1) s_{lj} w_j^{\mu_l - 1} (\sum_{j \in C_l} s_{lj} w_j^{\mu_l})^{\frac{\mu}{\mu_l}-2} \mu_l s_{lk} w_k^{\mu_l - 1}$
$= \mu \sum_{l=1}^{L} (\mu - \mu_l) s_{lj} s_{lk} (w_j w_k)^{\mu_l - 1} (\sum_{j \in C_l} s_{lj} w_j^{\mu_l})^{\frac{\mu}{\mu_l}-2} \leq 0$

From G we receive the choice probability of alternative j

$$\begin{aligned} P_j &= \frac{w_j \frac{\partial G}{\partial w_j}(w_1, ..., w_J)}{\mu G(w_1, ..., w_J)} \\ &= \frac{w_j \mu \sum_{l=1}^{L} s_{lj} w_j^{\mu_l - 1} (\sum_{j \in C_l} s_{lj} w_j^{\mu_l})^{\frac{\mu}{\mu_l}-1}}{\mu \sum_{r=1}^{L} (\sum_{k \in C_r} s_{rk} w_k^{\mu_r})^{\frac{\mu}{\mu_r}}} \\ &= \frac{w_j \sum_{l=1}^{L} s_{lj} w_j^{\mu_l} (\sum_{j \in C_l} s_{lj} w_j^{\mu_l})^{\frac{\mu}{\mu_l}-1} \frac{1}{w_j}}{\sum_{r=1}^{L} (\sum_{k \in C_r} s_{rk} w_k^{\mu_r})^{\frac{\mu}{\mu_r}}} \\ &= \frac{\sum_{l=1}^{L} s_{lj} w_j^{\mu_l} (\sum_{j \in C_l} s_{lj} w_j^{\mu_l})^{\frac{\mu}{\mu_l}-1}}{\sum_{r=1}^{L} (\sum_{k \in C_r} s_{rk} w_k^{\mu_r})^{\frac{\mu}{\mu_r}}} \end{aligned}$$

Replacing w_j by e^{V_j}, $j = 1, .., J$, yields the CNMNL model

$$P_j = \frac{\sum_{l=1}^{L} s_{lj} e^{\mu_l V_j} (\sum_{j \in C_l} s_{lj} e^{\mu_l V_j})^{\frac{\mu}{\mu_l}-1}}{\sum_{r=1}^{L} (\sum_{k \in C_r} s_{rk} e^{\mu_r V_k})^{\frac{\mu}{\mu_r}}} \tag{3.22}$$

Other popular GEV models are restricted versions of the CNMNL model. Wen and Koppelman's GNMNL model restricts the parameter $\mu = 1$. Based on the function

$$G(w_1, ..., w_J) = \sum_{l=1}^{L} (\sum_{j \in C_l} (s_{lj} w_j)^{\frac{1}{\mu_l}})^{\mu_l}$$

the choice probability becomes

$$P_j = \frac{\sum_{l=1}^{L} (s_{lj} w_j)^{\frac{1}{\mu_l}} (\sum_{j \in C_l} (s_{lj} w_j)^{\frac{1}{\mu_l}})^{\mu_l - 1}}{\sum_{r=1}^{L} (\sum_{k \in C_r} (s_{rk} w_k)^{\frac{1}{\mu_r}})^{\mu_r}} \tag{3.23}$$

By contrast, a preceding version of the CNMNL suggested by Vovsha constrains $\mu_l = 1$, $l = 1, ..., L$. Derived from the function G, which is given by

$$G(w_1, ..., w_J) = \sum_{l=1}^{L} (\sum_{j \in C_l} s_{lj} w_j)^{\mu}$$

Vovsha's CNMNL choice probability formula is expressed as

$$P_j = \frac{(\sum_{j \in C_l} s_{lj} w_j)^{\mu}}{\sum_{r=1}^{L} (\sum_{k \in C_r} s_{rk} w_k)^{\mu}} \tag{3.24}$$

And to end with, in order to obtain the NMNL model derived in (3.18), we restrict $s_{lj} = 1$, $j \in C_l$, $j = 1, ..., J$, $l = 1, ..., L$.

Chapter 4

Models with Integral Form

4.1 Introduction

4.2 Multinomial Probit Model

The search for flexible models that provide a better model fit than GEV models led to the emergence of the multinomial probit (MNP) model (Hausman and Wise, 1978; Daganzo, 1979). The MNP combines behavioral plausibility with flexibility by allowing for any patterns of heteroscedasticity and correlation that exist in the context of the multivariate normal distribution. However, the multifold integral over the choice probability formula cannot be solved analytically, and its evaluation requires simulation assistance.

4.2.1 Error Term Assumption

We assume here the usual random utility specification: $U_{ij} = V_{ij} + \epsilon_{ij}$, $i = 1, ..., n$, $j = 1, ..., J$. To get the MNP specification, one adds the theoretically appealing assumption that the error vector related to observation i is multivariate normally distributed with $J \times 1$ mean vector 0 (the null vector) and $J \times J$ variance-covariance matrix Σ:

$$\epsilon_i \sim MVN(0, \Sigma) \tag{4.1}$$

4.2.2 Choice Probabilities

Recall the $(J-1)$-dimensional integral representation of qualitative choice models:

$$P(y_{ij}=1) = \int \mathbf{1}(\epsilon_{ik} - \epsilon_{ij} < V_{ij} - V_{ik}, \forall k \neq j) f(\epsilon_{i,-j}) d\epsilon_{i,-j} \tag{4.2}$$

In a MNP, $\epsilon_{i,-j} \sim N(0, \Delta\Sigma)$, with $\Delta\Sigma$ denoting the $(J-1)\times(J-1)$ transformation variance-covariance matrix of the model introduced previously.

4.2.3 Identification of the Variance-Covariance Parameters

Following the approach of Bunch (1991), we will discuss identification and normalization issues associated with MNP models. In what follows, let us suppose the case where three alternatives are available for a particular individual, indexed by $j = 1, 2, 3$. From the assumptions above we know that the unidentified 3×3 variance-covariance matrix of the model comprises six unknown parameters:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ & \sigma_2^2 & \sigma_{23} \\ & & \sigma_3^2 \end{pmatrix}$$

To take differences in utilities, we define $\Delta U_{ik,-1} = U_{ik} - U_{i1}$, $k = 2, 3$, and get the 2×2 transformation variance-covariance matrix

$$\Delta\Sigma = \begin{pmatrix} \sigma_1^2 - 2\sigma_{12} + \sigma_2^2 & \sigma_1^2 - \sigma_{12} - \sigma_{13} + \sigma_{23} \\ & \sigma_1^2 - 2\sigma_{13} + \sigma_3^2 \end{pmatrix}$$

In $\Delta\Sigma$, there are still six unknown σ-parameters, but only three independent equations. The rank condition must hold, stating that only linear independent terms in $\Delta\Sigma$ can be estimated, minus one term normalized to set the scale of utility. Hence, for $J = 3$ alternatives in the choice set the maximum number of identified terms is 2 $((J(J-1)/2)-1)$. We normalize the first of the diagonal elements of the transformation matrix to one, and write

$$\Delta\Sigma^* = \begin{pmatrix} 1 & \omega_{23} \\ & \omega_{33} \end{pmatrix} \tag{4.3}$$

where $\omega_{23} = \frac{\sigma_{23} - \sigma_{12} - \sigma_{13} + \sigma_1^2}{\sigma_1^2 - 2\sigma_{12} + \sigma_2^2}$ and $\omega_{33} = \frac{\sigma_1^2 - 2\sigma_{13} + \sigma_3^2}{\sigma_1^2 - 2\sigma_{12} + \sigma_2^2}$ are the terms to be estimated. Note that normalizing one of the σ-terms is not sufficient to set the scale of utility. If $\sigma_1^2 = 1$, for example, infinite many values for $1 - 2\sigma_{12} + \sigma_2^2$ provide equivalent models.

4.2.4 Factor-Analytic Form

To estimate ω_{23} and ω_{33}, we impose the following factor-analytic structure on the error terms suggested by Hausman and Wise:

$$\Delta U_{i2,-1} = V_{i2} - V_{i1} + c_{11}\eta_{i1}$$

$$\Delta U_{i3,-1} = V_{i3} - V_{i1} + c_{21}\eta_{i1} + c_{22}\eta_{i2}$$

Here, we assume $\eta_i \sim MVN(0, I)$, where I is the 2×2 identity matrix. c_{11}, c_{21} and c_{22} are elements of a 2×2 matrix C denoted Cholesky factorization and satisfying $CC' = \Delta\Sigma$. For estimation, we normalize $c_{11} = 1$ to set the scale of utility. Accordingly to this, the factor-analytic expression of (4.3) is

$$\Delta\Sigma^* = \begin{pmatrix} 1 & c_{21} \\ & c_{21}^2 + c_{22}^2 \end{pmatrix} \tag{4.4}$$

The complexity of simulation based estimation increases dramatically with the number of outcomes in the choice set. An increase in the number of alternatives from J to $J+1$ augments the number of unknown parameters by $J+1$ and the maximal number of identified parameters by $J-1$. Hence, If the number of outcomes is large and simulation assisted estimation is extremely time-consuming, it may be reasonable to impose an a priori structure, for example the heteroscedastic structure, on the original variance-covariance matrix.

4.2.5 Maximum Simulated Likelihood Estimation

The analytically intractable integral over the multivariate normal density makes exact ML estimation of the MNP model infeasible. Rather, we apply MSL estimation that involves probabilities that are approximated through simulation. The basic idea behind simulation is to replace the multifold response probability integral with easy to compute probability simulators. Once the MNP probabilities are made numerical, the log-likelihood function can be calculated for a each parameter vector.

Though commonly used MNP simulators are unbiased for probabilities appearing linearly across observations in an estimation procedure, MSL estimation is biased because of the logarithmic transformation of the likelihood function. Nevertheless, the need for simulation assistance does not necessarily affect the estimation results negatively. As already explained in Chapter 2, MSL can approximate ML when the number of simulation draws is sufficient large.

4.2.6 GHK Simulation

The GHK simulation has been found to be particularly appropriate to approximate MNP choice probabilities. To illustrate the simulation procedure, we calculate the probability of choosing the first alternative of a trinomial choice set. As customary, we start with taking utility differences and imposing factor-analytic error term structures: $\Delta U_{i2,-1} = V_{i2} - V_{i1} + c_{11}\eta_{i1}$ and $\Delta U_{i3,-1} = V_{i3} - V_{i1} + c_{21}\eta_{i1} + c_{22}\eta_{i2}$. Then the probability that observation i chooses alternative 1 is given by

$$\begin{aligned} P(y_{i1} = 1) &= Prob\{\Delta U_{ik,-1} < 0, k = 2, 3\} \\ &= Prob\{V_{i2} - V_{i1} + c_{11}\eta_{i1} < 0, V_{i3} - V_{i1} + c_{21}\eta_{i1} + c_{22}\eta_{i2} < 0\} \end{aligned}$$

which can be written as the product of the marginal and the conditional probability

$$\begin{aligned} P(y_{i1} = 1) &= Prob\{V_{i2} - V_{i1} + c_{11}\eta_{i1} < 0\} \\ &\quad *Prob\{V_{i3} - V_{i1} + c_{21}\eta_{i1} + c_{22}\eta_{i2} < 0 | V_{i2} - V_{i1} + c_{11}\eta_{i1} < 0\} \end{aligned}$$

In terms of cumulative densities, we obtain

$$\begin{aligned} P(y_{i1} = 1) &= Prob\{\eta_{i1} < \frac{V_{i2} - V_{i1}}{c_{11}}\} * Prob\{\eta_{i2} < \frac{V_{i3} - V_{i1} + c_{21}\eta_{i1}}{c_{22}} | \eta_{i1} < \frac{V_{i2} - V_{i1}}{c_{11}}\} \\ &= \Phi(-\frac{V_{i2} - V_{i1}}{c_{11}}) * \int_{\eta_{i1}=-\infty}^{\frac{V_{i2}-V_{i1}}{c_{11}}} \Phi(\frac{V_{i3} - V_{i1} + c_{21}\eta_{i1}}{c_{22}}) f(\eta_{i1}) d\eta_{i1}\} \end{aligned}$$

The first term is a cumulative standard normal distribution that can be evaluated by computer software library routines for a given set of parameters (c_{11}, c_{21}, c_{22}). The second term is an integral which must be simulated using random draws from a truncated univariate normal distribution. The same is true for the GHK simulation of the probabilities $P(y_{i2} = 1)$ and $P(y_{i3} = 1)$.

4.2.7 Binary Probit Model

Consider decision-maker i's binary decision problem, represented by the random utilities $U_{i1} = V_{i1} + \epsilon_{i1}$ and $U_{i2} = V_{i2} + \epsilon_{i2}$. Unlike MNP, ϵ_i is assumed to arise from an *independent* bivariate normal distribution. The binary probit (BP) choice probability of i are as follows:

$$\begin{aligned} P(y_{i1} = 1) &= Prob\{\epsilon_{i2} - \epsilon_{i1} < V_{i1} - V_{i2}\} \\ &= F\epsilon_{i2} - \epsilon_{i1}(V_{i1} - V_{i2}) \end{aligned}$$

and $P(y_{i2} = 1) = 1 - P(y_{i1} = 1)$

F is the cumulative distribution function of the normal distribution $N(0, \sigma_1^2 + \sigma_2^2)$. Since the sum of the variances $\sigma_1^2 + \sigma_2^2$ is not identified, we can normalize the single variance elements by setting $\sigma_1^2 = 0.5$ and $\sigma_2^2 = 0.5$, thus $\epsilon_{i2} - \epsilon_{i1} \sim N(0, 1)$. And so:

$$P(y_{i1} = 1) = \Phi(V_{i1} - V_{i2}) \tag{4.5}$$

where Φ is the cumulative density of a bivariate standard normal variable.

4.3 Mixed Multinomial Logit Models

Based on convenient error term partitioning, mixed MNL (MMNL) models are very general models that can approximate any behavior consistent with RUM as closely as desired (McFadden and Train, 2000). One portion of the error term is IID EV1, which keeps the basic model a MNL, while the mixing portion can follow any distribution. When the mixing distribution is multivariate normal, the mixed MNL model is formally a MNP including an IID EV1 error term and thus also denoted as ”MNP with logit kernel”.

This idea of MMNL models as extensions of the MNL is not new. Similar models have already been suggested by Cardell and Dunbar (1980) and Boyd and Melman (1980). Yet only recent improvements in computer capacity and speed have helped to get the idea accepted.

4.3.1 Error Components Multinomial Logit Model

The error components MNL (ECMNL) model comprises the same error term as the MNL, and then adds additive error components that can be heteroscedastic or correlated between alternatives or both to provide realistic substitution patterns between the outcomes.

4.3.2 Assumptions on the Error Terms

Consider the random utility representation $U_{ij} = V_{ij} + \xi_{ij} + \epsilon_{ij}$, $i = 1, ..., n$, $j = 1, ..., J$.

Without loss of generality, we assume as follows:

1. ϵ_i is a $J \times 1$ vector of IID EV1 distributed error terms with location parameters $\tau_j = 0$ for all $j \in C$ and scale parameter μ.

2. ξ_i is a $J \times 1$ vector of random terms called error components, which can represent any distribution.

3. ξ_i and ϵ_i are independent.

Different patterns of correlation and heteroskedasticity are obtained through the appropriate specification of ξ_i. In econometric literature, ξ_i is usually multivariate normally distributed:

$$\xi_i \sim MVN(0, \Omega) \tag{4.6}$$

4.3.3 Choice Probabilities

We can exploit the fact that, given ξ_i, the IID EV1 error portion can be integrated analytically, and the probability that an individual i chooses alternative j is a MNL

$$P(y_{ij} = 1)|\xi_i = \frac{e^{\mu V_{ij} + \xi_{ij}}}{\sum_{k=1}^{J} e^{\mu V_{ik} + \xi_{ik}}} \tag{4.7}$$

However, ξ_i is unknown, and the unconditional ECMNL choice probability is the integral of the conditional probability over all possible values of the mixing (the multivariate normal) distribution

$$P(y_{ij} = 1) = \int \frac{e^{\mu V_{ij} + \xi_{ij}}}{\sum_{k=1}^{J} e^{\mu V_{ik} + \xi_{ik}}} \; f(\xi_i) \; d\xi_i \tag{4.8}$$

4.3.4 Identification of the Variance-Covariance Matrix

Recent contributions of Walker (2002), Chiou and Walker (2006) and Walker et al. (2006) disagree with the prevalent belief that identification (and normalization) of ECMNL and MNP models is identical. In fact, the logit kernel characteristic of ECMNL models that is not present in analogous MNP models entails one more term to be estimated in ECMNL than in MNP models.

Analytically, the $J \times J$ variance-covariance matrix of the ECMNL model is composed of two portions

$$\Sigma = \Omega + \frac{\pi^2}{6\mu^2} I$$

In Σ, Ω is the $J \times J$ variance-covariance matrix of ξ_i, and $\frac{\pi^2}{6\mu^2} I$ is the $J \times J$ variance-covariance matrix of ϵ_i. Considered apart there are at most $(J(J-1)/2) - 1$ identified terms related to Ω and zero identified terms related to $\frac{\pi^2}{6\mu^2} I$. Nevertheless, Σ contains $J(J-1)/2$ identified terms, one more than in a MNP.

To make this point clear, suppose again a trinomial choice set accompanied with the following 3×3 variance-covariance matrix:

$$\Sigma = \begin{pmatrix} \frac{\pi^2}{6\mu^2} + \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ & \frac{\pi^2}{6\mu^2} + \sigma_2^2 & \sigma_{23} \\ & & \frac{\pi^2}{6\mu^2} + \sigma_3^2 \end{pmatrix}$$

After taking differences in utilities, $\Delta U_{ik,-1} = U_{i2} - U_{i1}$, $k = 2, 3$, we normalize $\mu = 1$, and write

$$\Delta\Sigma^* = \begin{pmatrix} \frac{\pi^2}{3} + \omega_{22} & \frac{\pi^2}{6} + \omega_{23} \\ & \frac{\pi^2}{3} + \omega_{33} \end{pmatrix} \tag{4.9}$$

where $\omega_{22} = \sigma_1^2 - 2\sigma_{12} + \sigma_2^2$, $\omega_{23} = \sigma_1^2 - \sigma_{12} - \sigma_{13} + \sigma_{23}$, and $\omega_{33} = \sigma_1^2 - 2\sigma_{13} + \sigma_3^2$ are identified terms. Unlike MNP, there is no need to normalize one ω-term since the restriction $\mu = 1$ has already fixed the scale of utility, relative to which the ω-terms can be estimated. For estimation, the factor-analytic structure is implemented, as described in the MNP case.

4.3.5 Random Parameter Multinomial Logit Models

Even after developing the best systematic specification for response heterogeneity it is very unlikely that we have information on all determinants of choice, so that there may still be unobserved taste variation in the population. The random parameter MNL (RPMNL) model generalizes the MNL model by allowing the parameters on observed attributes of the alternatives to vary randomly over observations rather than to be fixed.

4.3.6 Assumptions on the Random Terms

Formally, random utility is specified as $U_{ij} = c_j + x'_{ij}\beta + z'_i\alpha_j + q'_{ij}\gamma_i + \epsilon_{ij}$, $i = 1, ..., n$, $j = 1, ..., J$. The assumption bundle can be summarized in the following way:

1. ϵ_i is a $J \times 1$ vector of IID EV1 distributed error terms with location parameters $\tau_j = 0$ for all $j = 1, ..., J$, and scale parameter μ.
2. γ_i is a $M \times 1$ vector of population parameters related to the vector of alternative-specific attributes q_{ij}, varying randomly (unobserved) across the observations $i = 1, ..., n$.
3. γ_i and ϵ_i are independent.

We are free to assign the alternative-specific attributes to the vector x_{ij}, which is associated with the vector of fixed taste parameters β, or the vector q_{ij}, which is related to the vector of random parameters γ_i. But note that no attribute can appear in x_{ij} and q_{ij} at the same time.

A theoretically limitless number of distributions can be used for random parameters. Nevertheless, γ_i is usually specified to arise from a multivariate normal distribution with $M \times 1$ mean vector γ and $M \times M$ variance-covariance matrix Ψ

$$\gamma_i \sim MVN(\gamma, \Psi) \tag{4.10}$$

4.3.7 Discussion of Normal Distribution Assumption

Assumption (4.10) is accompanied by the question whether the use of the normal distribution is consistent with the hypothesis of rational consumer behavior. For example, we usually believe that the demand for a good is a negative function of the price of this good. Yet when applying unbounded mixing distributions like the normal distribution we *assume* that some consumers attach positive weight on higher prices. Then it is impossible to decide whether the observed non-zero probability of irrational behavior is revealed by the data or is an artifact of the unbounded nature of the distribution[5].

To avoid this difficulty, random parameters can be specified to follow other distributions, such as the uniform and triangular, which are bilaterally bounded, or the log-normal, which is unilaterally bounded. However, the main drawback of these distributions is their lack of behavioral plausibility.

4.3.8 Choice Probabilities

The computational procedure for obtaining RPMNL choice probabilities is the same as previously. Provided that the random parameters γ_i are known, the integration over IID error terms leads to the following closed-form MNL probability:

$$P(y_i = j|\gamma_i) = \frac{exp(c_j + x'_{ij}\beta + z'_i\alpha_j + q'_{ij}\gamma_i)}{\sum_{k=1}^{J} exp(c_k + x'_{ik}\beta + z'_i\alpha_k + q'_{ik}\gamma_i)} \tag{4.11}$$

[5] See Hess et al. (2005) for a detailed discussion of this problem in the context of evaluating the value of travel time savings in transportation economics.

The unconditional probability of interest is obtained by the integration of the conditional choice probability over the joint distribution of unobserved random parameters, denoted as $g(\gamma_i)$:

$$P(y_{ij} = 1) = \int \frac{exp(c_j + x'_{ij}\beta + z'_i\alpha_j + q'_{ij}\gamma_i)}{\sum_{k=1}^{J} exp(c_k + x'_{ik}\beta + z'_i\alpha_k + q'_{ik}\gamma_i)} g(\gamma_i)\ d\gamma_i \tag{4.12}$$

4.3.9 Identification of γ and Ψ

In general, it is not difficult to estimate the mean vector γ and the variance-covariance matrix Ψ of the normal distribution (4.10). To make this point clear, rewrite random utilities as $U_{ij} = c_j + x'_{ij}\beta + z'_i\alpha_j + q'_{ij}\gamma + q'_{ij}C\eta_i + \epsilon_{ij}$, where C is a Cholesky factor satisfying $CC' = \Psi$, $\eta_i \sim N(0, 1)$. Then $q'_{ij}C\eta_i$ is the normally distributed stochastic deviation representing the individual's tastes relative to the average tastes in the population. As becomes clear, the rule that identified parameters must capture differences between alternatives is fulfilled by the mean vector γ as well as the Cholesky factor C.

4.3.10 Variance-Covariance Matrix

As a by-product, unobserved taste heterogeneity *can* produce unobserved heteroscedasticity and correlation between utilities, expressed by the $J \times J$ variance-covariance matrix of the model

$$\Sigma = q'_i\Psi q_i + \frac{\pi^2}{6\mu^2}I$$

In Σ, the first portion is related to $q_i\gamma_i$, and the second portion is associated with the vector of IID EV1 errors. Since the vetor q_i can include positive as well as negative terms, the portion $q'_i\Psi q_i$ can be zero even if Ψ is not the null matrix.

4.3.11 Equivalence of Random Parameter and Error Components Multinomial Logit Models

When are ECMNL and RPMNL formally equivalent? To examine this question, rewrite the ECMNL model as

$$U_i = \alpha_0 + x'_i\beta + z'_i\alpha + C_\Omega\eta_i + \epsilon_i \tag{4.13}$$

where C_Ω satisfies $C_\Omega C'_\Omega = \Omega$, and compare it with the RPMNL model from before

$$U_i = \alpha_0 + x'_i\beta + z'_i\alpha + q'_i\gamma + q'_iC_\Psi\eta_i + \epsilon_i \tag{4.14}$$

where C_Ψ fulfills $C'_\Psi C_\Psi = \Psi$. Hence, the EMNL and the RPMNL model are equal in the special case when the vector x_i of the RPMNL model is the null vector and $q'_i C_\Psi = C_\Omega$.

4.3.12 Maximum Simulated Likelihood Estimation

Whenever analytically unsolvable high-dimensional probability integrals are involved as in (4.8) or (4.12), the log-likelihood function for the sample comprises simulated probabilities. Due to the logarithmic transformation, the simulated objective function is biased even if probability simulation is unbiased. In order to minimize this bias, it is crucial to simulate the probabilities with good precision by using large PRMC sequences during the approximation process or by using variance reduction techniques such as QRMC drawing. Experiments by Ben-Akiva and Bolduc (1996) indicated that simulation using 250 PRMC draws or 100 QRMC draws such as antithetic draws is large enough to produce negligible bias.

4.3.13 Frequency Simulation

MMNL choice probabilities are typically approximated through frequency (accept-reject) simulation, which can be described as follows:

1. For a given set of population parameters, make $d = 1, ..., D$ MC draws from the mixing density $f(\xi_i)$ or $g(\gamma_i)$, respectively.
2. For each draw d, calculate the conditional MMNL choice probabilities presented in (4.7) or (4.11), respectively.
3. The average of the conditional probabilities is taken as the simulated choice probability of an alternative j: $\check{P}(y_{ij} = 1) = \frac{1}{D} \sum_{d=1}^{D} P(y_{ij} = 1)|\epsilon_i^d$

By construction, $\check{P}(y_{ij} = 1)$ is an unbiased estimator of $P(y_{ij} = 1)$ for any number of draws D, and its variance decreases as D increases. It is strictly positive for any D, so that $log\check{P}(y_{ij} = 1)$ is always defined, which is important when using $\check{P}(y_{ij} = 1)$ in the log-likelihood function.

In general, the occurrence of IID error variables along with multivariate normal error terms is irrelevant for choice. In MMNL models, however, the multifold integral over these variables can be calculated exactly. Then the crude frequency simulator becomes a smooth logit-kernel frequency simulator, which is the average of a set of MNL probabilities and which is unbiased since the MNL kernel function is a "natural" smooth function resulting from error partitioning.

Part II

Empirical Findings

Chapter 5

Data Base and Samples

5.1 Introduction

We utilize data obtained from the Swiss census 2000 as the primary data source. In this chapter, we describe the census record and the methodology applied to derive missing variables endogenously from various data sources. Furthermore, we define the relevant commuter population and the scope of the study field, and give summaries of the final samples used to realize the empirical investigations.

5.2 Swiss Census Record 2000

5.2.1 Aim and Purpose

The Swiss census is a survey that is administered decennially by the Swiss Federal Statistical Office (SFSO) with the aim to capture the entire population and to trace demographic, spatial, social and economic developments on Switzerland's national territory. Participation in the census is incumbent on everybody domiciled fiscally and legally in Switzerland.

The Swiss census was implemented for the first time in 1850, and since then there have been several innovations in data collection. Alongside the original function as a pure inhabitant count, the census was modified and extended repeatedly to a survey unveiling a variety of supplementary information of the respondents.

5.2.2 Questionnaire-Based Survey

The methodology of the latest survey was based on three questionnaires. The personal questionnaire covered a total of 21 subject areas, involving questions on characteristics of the respondent such as (among others) birthday, sex, marital status, highest educational level, questions on geographic characteristics such as place of residence and place of work, and questions on the behavior on the journey to work such as means of transportation, average total travel time per trip, and trip frequency per day.

The household questionnaire was a means to gather information about the respondent's household and to identify people living in the same accommodation. Furthermore, homeowners had to complete a questionnaire that provided information about the respondent's living space. It contained 19 questions on attributes related to the residential house, including (among others) exact address, ownership, number of flats, number of rooms, area, and monthly rent a flat.

5.2.3 Dataset and Data Quality

In 2000, the data of altogether 7,288,010 people living in Switzerland were gathered by December 5th. Return run controls by the SFSO showed that only 0.013 percent of the population failed answering.

The quality of the answers was also checked. Incomplete or carelessly filled out questionnaires were sent back to the respondents or were completed as far as possible by the local authorities and the SFSO. As a result, variables that could be verified by local authorities such as birth date, sex, marital status, and so on appear without gap. After evaluating the mass of standardized questionnaires, the SFSO made full and partial data records available for research purposes.

Overall, the census exhibits nice properties that facilitate research. Compared to pure travel surveys, it provides a vast number of additional information about the respondents and their personal and vocational environment. On the other hand, however, we must determine incomplete or missing time and money cost variables endogenously from various data sources (see Chapter 5.6.2).

5.3 Commuter Definition

In home-to-work-to-home trip models, the commuter is the central decision-maker unit. Henceforth, a commuter is defined as a person (1) who works at least one hour per week, (2) who is of age (i.e. who is at least eighteen years old), and (3) who does not live and work in the same commune.

(1) is the condition that defines a working person according to the SFSO. (2) is imposed to ensure that each commuter is able to acquire a license to drive if required. However, the data do not reveal whether the respondent actually possessed one or not. And finally, (3) serves to distinguish working people who leave their municipality in order to go to work from those who do not.

5.4 Study Field

We restrict our investigations to commuters who live and work in the canton of Zurich. Thus, work trips that involve crossing inter-cantonal borders are not considered. An essential reason for narrowing the study field is that the missing trip cost variable is relatively easy to determine for travel within the canton of Zurich. Since public transport is pooled in the Zurich Traffic Association (Zuercher Verkehrsverbund), rail and bus fares are subject to a unified tariff system.

The canton of Zurich comprises a total of 171 political communes, including the cities Zurich and Winterthur. In accordance to the definition above, we consider that a worker commutes when he lives in one of these 171 communes and works in another. The relevant commuter population can be identified because the data provide the exact address of both the respondent's residence and work place. Pursuant to the census, 314,724 of 1,267,478 people living in the canton of Zurich belonged to the condition group.

5.5 Choice Variable

5.5.1 Travel Means

The personal questionnaire gave a selection of eleven transport means, whereas any combination was allowed. In the final data record, the SFSO categorized the answers into ten main categories with subcategories organized hierarchically, in compliance with the following rules: Private travel means were subordinated to public transport means. Within both groups, slow travel means were subordinated to fast travel means.

Finally, the ten main categories include rail, regional transport (regional bus), urban transport (urban bus or tramway), other public transport, company bus, car, motorbike, moped, bike, and walking. Due to multiple mentions, each category is divided into subcategories. For instance, the main category rail involves rail alone, rail and regional transport, rail and urban transport, rail and another public transport, rail and company bus, rail and car, rail and motorbike, rail and moped, and rail and bike[6].

To a certain degree, categorization made by the SFSO was arbitrary. The rule to generally subordinate private to public transport, in particular the car to the bus and to other public transportation means (except rail), seems problematical. For example, commuters who ticked *car* as well as *bus* can be found in the category bus (subcategory bus and car), although it is improbable that someone drives the car to the bus stop to take the bus to work. Rather, this person is more likely a car driver, going by bus from the parking lot to work.

As a result, the shares of bus, other public transport and company bus tend to be overrated, at the expense of the automobile. For this reason, we recode three subcategories in order to reestablish the actual market shares, including bus and car, other public transport and car, and company bus and car.

Table 5.1: Commuters' Travel Mode Choices

	Frequency	Market Share (in %)
Rail	93,590	29.74
Bus	15,654	4.97
Other public transport	408	0.13
Company bus	1,936	0.62
Car	182,053	57.85
Motorbike	4,041	1.28
Bike	4,766	1.51
On foot	1,571	0.50
No statement	10,705	3.40
All	314,724	100

Notes: Own calculations, based on Swiss census 2000. The uncorrected numbers are: bus 6.39 percent (+1.42 percent), other public transport 0.31 percent (+0.18 percent), company bus 0.74 percent (+0.12 percent), car 56.13 percent (−1.72 percent).

5.5.2 Distribution in the Canton of Zurich

Table 5.1 unveils the corrected frequencies and percentage market shares in the canton of Zurich[7]. For simplicity, the categories regional transport (regional bus) and urban transport (urban bus and tram) are summarized henceforth in the category *bus*.

It can be seen from the table that the majority of commuters traveled by car to work, followed by

[6] Walking appears in the rail alone category.

[7] The uncorrected numbers can be found in the table notes.

rail and bus. Since these leading transportation means accounted for approximately 96 percent of all statements (not counting 3.4 percent missing values), for practical reasons further examinations forgo the other travel means.

5.6 Explanatory Variables

5.6.1 Definitions

Table 5.2 gives the definitions of the predictors which are hypothesized to affect mode choice behavior significantly. These include level-of-service attributes, socio-demographic characteristics of the decision maker, and the household's monthly rent of the dwelling. With the exception of rent, all variables listed in the table are measured at the individual level. Further, note that male, married, kids, foreigner, part-time, medium education, and high education are zero-one dummy variables.

Table 5.2: Definitions of Explanatory Variables

Variable	Description
Out-of-vehicle time	Time spent out of the main vehicle, one way, measured in minutes
On-vehicle time	Time spent in the main vehicle, one way, measured in minutes
Total cost	Total travel cost, one way, measured in Swiss francs
Age	Age in years
Male	1 if a male, 0 otherwise
Married	1 if married, 0 otherwise
Children	1 if person brings up at least one child (aged below eighteen), 0 otherwise
Foreigner	1 if a foreigner, 0 if a Swiss
Part-time	1 if working less than 42 hours per week, 0 otherwise
Medium education	1 if highest completed education is secondary school or apprenticeship, 0 if highest completed education is at best compulsory school or at least college of higher education or university
High education	1 if highest completed education is technical college or university, 0 if highest completed education is at best secondary school or apprenticeship
Rent	Monthly rent, measured in 1000 Swiss francs

5.6.2 Level-of-Service Attributes

The following level-of-service variables are explored: on-vehicle time, out-of-vehicle time, and total cost of the trip. From an economic point of view, these attributes represent the price of commuting. Longer and costlier travels shorten leisure and reduce consumption, respectively.

The variable total cost includes the whole money cost incurring during the trip. Total trip time can be subdivided into an on-vehicle time and out-of-vehicle time portion. As the name says, on-vehicle time measures the minutes spent on the main vehicle. Out-of-vehicle time involves activities such as walking, waiting and so on. For car trips out-of-vehicle time is assumed to be zero. Hence, by subdividing total time, we learn whether rail and bus commuters evaluate on-vehicle and out-of-vehicle time differently or not.

Unfortunately, the census exhibits several drawbacks, including data imperfections. Neither of the mentioned attributes can be found directly in the census, so that we must complete the data base. In the following, we describe the way we derive the missing variables endogenously from various data sources.

Endogenous Determination of Out-of-Vehicle and On-Vehicle Time

In the personal questionnaire, each working person was asked to state the average total travel time (exact to the minute) for the mode alternative which was usually chosen on the trip to work. Yet the following is missing:

1. Distinction between on-vehicle and out-of-vehicle time scores

2. Indication of hypothetic travel times for travel modes not selected

Rail and bus out-of-vehicle time scores can be determined independently from the total travel time scores in the census. For bus, we assume that out-of-vehicle time is composed of three portions, namely the time required for going from the residence to the next bus stop, the waiting time for the bus and the time required for reaching the work place after deboarding the bus. We assume an average walking speed of 12 minutes per kilometer (or 5 kilometers per hour) and an average waiting time at the bus stop of 3 minutes.

Rail out-of-vehicle time scores are also composed of several time sequences. Assumptions regarding the time sequence from leaving the residence to boarding the train can be summarized as follows:

- If the distance between residence and train station is shorter than 0.5 km, commuters walk (average speed: 12 minutes per km) and wait an average of 3 minutes for the train at the train station.

- If the distance between residence and train station is more than 0.5 km, however, the distance from residence to the next bus stop is shorter than 0.5 km, commuters walk to the bus stop (average walking speed: 12 minutes per km), wait an average of 3 minutes for the bus, ride the bus to the train station (average speed: 4 minutes per km) and wait an average of 3 minutes for the train.

- If both distances are longer than 0.5 km, commuters drive their cars to the train station (average speed: 2 minutes per km), require an average of 3 minutes for parking the car, and wait an average of another 3 minutes for the train.

For the out-of-vehicle time from the train station to the work place, we assume as follows:

- If the distance from the train station to the work place is shorter than 0.5 km, commuters walk (average speed: 12 minutes per km).
- If the distance from the train station to the work place is longer than 0.5 km, however, the distance from the bus stop to the work place is shorter than 0.5 km, commuters ride the bus[8] (average speed: 4 minutes per km) and walk from the bus stop to the work place (average speed: 12 minutes per km)
- If both distances are longer than 0.5 km, commuters were collected by a car or a company bus (average speed: 2 minutes per km).

In order to calculate the corresponding distances (in km) we compare the census data containing the exact coordinates of residence and work place with the data of the Zurich Traffic Association containing the exact coordinates of all 175 train stations and 1832 bus stops in the canton of Zurich. Based on the coordinates, the corresponding airline distances are calculated and multiplied by 1.4 to approximate the walking and driving distances, respectively.

The rail and bus on-vehicle time is calculated from the difference between the total travel time and the out-of-vehicle time. For cars, the out-of-vehicle time is assumed to be zero, and the commuter's on-vehicle time is identical to the total travel time score in the census. For an actually selected travel means, the on-vehicle time is therefore simply calculated according to the above directions. Unfortunately, the data do not give direct information on the on-vehicle time of alternatives not selected. For example, individuals taking the car were not asked about the total time for the rail and the bus alternatives.

[8] At each train station in the canton Zurich there is a bus stop.

To fill out the on-vehicle time on a certain route for the alternatives not selected, our procedure is to use the average on-vehicle time score of these commuters who effectively used the certain traffic means on the certain route. As an average the median is applied due to its insensibility to outliers. Numerous examinations of various randomly selected distances showed that the endogenously calculated on-vehicle time values (and therefore also the out-of-vehicle time values) are plausible.

Endogenous Determination of Trip Cost

The procedures to compute the money cost of commuting are subjected to the means of transportation, which is considered. For rail and bus journeys, costs are calculated using a zone table and a price list of the year 2000 provided by the Zurich Traffic Association. The number of zones through which a commuter travels on the way from home to the work place determines the fare. The costs per journey are computed by assuming that commuters choose that second-class ticket which is cheapest calculated over the whole year. In most cases it was the monthly or annual ticket.

The costs for driving a car are the product of the distance between home and work place (airline distance in km multiplied by 1.4) and the average car costs per km, which are based on calculations of the Touring Club of Switzerland, the largest Swiss automobile club. It is assumed that all commuters drive an average car valued at approximately 28,000 Swiss francs. Further, car-parking cost is presumed to amount to 5 Swiss francs per trip day.

5.6.3 Socio-demographic Characteristics

The Swiss census is unique in the sense that rich information on personal characteristics is available, including age, gender, education, and so forth. These variables are introduced into travel mode choice models to control choice heterogeneity among different groups in the population.

The definitions in Table 5.2 require a few additional explanations. According to the table, married equals one if the commuter is married, and zero otherwise. Otherwise means that the person is single, divorced or widowed. The variable kids takes the value one if the person brings up at least one child below eighteen, and zero otherwise. Here, otherwise means that the person does not have children or has only children who are older than eighteen. The variable part-time defines every work relationship of at least one but less than 42 working hours per week. Persons who work less than one hour are not considered according to the before mentioned commuter definition. If the person works 42 or more hours, part-time takes the value zero.

The dummy variables middle education and high education also require further explanations. Both are defined relative to the reference category low education. The variable low education is one if the person's highest education is at best compulsory school, which usually involves nine years of schooling. If compulsory school is followed by an apprenticeship (between two and four years) or if the respondent completed a secondary school, education is medium (or "average"). High education takes the value one if the person graduated from university or technical college.

5.6.4 Monthly Rent

The variable rent measures the household's monthly cost of housing, measured in 1000 Swiss francs. Since it is missing for homeowners, we must determine the values endogenously from the data too. For each observation, the rent is predicted using a multiple linear regression model, which comprises the following covariates: The number of rooms of the dwelling, the area in square meters, dummies representing the periods of construction, dummies representing the periods of renovation, and region dummies. In doing so, we classify the period of construction and renovation into nine and five age-group classes, respectively, as in the census record. Moreover, we divide the canton of Zurich into seven regions, including the cities Zurich and Winterthur, four agglomeration belts, and the rural area, as defined by the SFSO.

5.7 Final Samples

5.7.1 Short-Term Sample

The final sample used to study commuters' short-term choices from the choice set {rail,bus,car} counts a total of 48,074 observations. Compared to Table 5.1, it has suffered a loss of 83.5 percent of the observations due to the endogenous computation of level-of-service attributes. On the one hand, the coordinates of home and work place are available only for around a third of the population. On the other hand, on-vehicle time scores for the alternatives not selected on a certain route are only determinable when there were people selecting them. For example, when there was nobody riding the bus on a certain route, the median bus time score is missing, and all observations traveling on this route are eliminated.

Table 5.3: Distribution of Short-Term Sample

	Rail	Bus	Car	All
Frequency	24,048	3,747	20,279	48,074
Sample Share (in %)	50.02	7.80	42.18	100
Out-of-vehicle time	20.06	10.42	0.0	10.16
	(5.65)	(6.12)	(0.00)	(9.50)
On-vehicle time	22.40	26.37	25.89	24.88
	(10.04)	(9.28)	(7.87)	(9.28)
Total cost	2.99	2.35	14.06	6.47
	(2.02)	(1.17)	(4.76)	(6.19)
Age	37.33	38.14	41.61	39.20
	(12.07)	(12.45)	(11.76)	(12.15)
Male	0.50	0.41	0.67	0.56
	(0.50)	(0.49)	(0.47)	(0.50)
Part-time	0.48	0.48	0.37	0.43
	(0.50)	(0.50)	(0.48)	(0.50)
Male×part-time	0.18	0.14	0.18	0.17
	(0.38)	(0.35)	(0.38)	(0.38)
Married	0.44	0.47	0.56	0.49
	(0.50)	(0.50)	(0.50)	(0.50)
Children	0.37	0.41	0.52	0.44
	(0.48)	(0.49)	(0.50)	(0.50)
Married×children	0.30	0.34	0.43	0.36
	(0.46)	(0.47)	(0.50)	(0.48)
Foreigner	0.16	0.27	0.20	0.19
	(0.37)	(0.44)	(0.40)	(0.39)
Low education	0.12	0.23	0.11	0.12
	(0.32)	(0.42)	(0.31)	(0.33)
Foreigner×low education	0.04	0.13	0.06	0.06
	(0.20)	(0.34)	(0.24)	(0.24)
Medium education	0.56	0.50	0.55	0.55
	(0.50)	(0.50)	(0.50)	(0.50)
Foreigner×medium education	0.07	0.07	0.08	0.07
	(0.26)	(0.26)	(0.27)	(0.26)
High education	0.32	0.27	0.34	0.33
	(0.47)	(0.44)	(0.48)	(0.47)
Foreigner×high education	0.05	0.06	0.05	0.05
	(0.22)	(0.24)	(0.22)	(0.22)

Notes: Calculations are based 48,074 observations. Standard deviations in parentheses.

5.7.2 Distribution of Short-Term Sample

The second line of Table 5.3 gives the distribution of the sample observations across the three work-trip modes: 50.02 percent rail, 7.80 percent bus, and 42.18 percent car. This distribution strongly differs from the market shares encountered in Table 5.1, which unveiled 32.1 percent rail, 5.4 percent bus, and 62.5 percent car. Train and bus riders are obviously overrepresented, car drivers, however, are underrepresented. The reason for this is twofold. First, in the data the probability of available home residence coordinates for persons living in the city or the suburbs is more likely than for persons living in the country. Second, we only consider commuters who faced the whole set of alternatives on their work trip. As mentioned above, the observations are eliminated when at least one alternative was not available (and thus at least one median time score is missing).

Beneath, the table displays summary statistics of level-of-service attributes and individual socio-demographic characteristics. The last row gives mean values of the whole sample. For example, the average journey to work lasts approximately 35 minutes, including approximately 10 minutes out-of-vehicle time and 25 minutes on-vehicle time, and costs a little bit less than 6.5 Swiss francs.

Dummy variables are often related to each other, and interaction terms are a means to visualize these associations. First consider male and part-time. As the values in the last row reveal, there are 56 percent males (females: 44 percent), 43 percent part-time workers (full-time: 57 percent), and 17 percent are part-time working males. From this we learn that in the sample there are 17 percent males working part-time, 39 percent males working full-time, 26 percent females working part-time, and 31 percent females working full-time. From the fact that 49 percent are married, 44 percent have children, and 36 percent are married with children we discover that 36 percent are married with children, 13 percent married without any children, 8 percent unmarried with children, and 43 percent unmarried without any children. And finally, the table unveils that 6 percent of the sample are foreigners with low, 7 percent foreigners with middle, and 5 percent foreigners with high education, while 6 percent of the observations are Swiss with low, 48 percent Swiss with middle, and 28 percent Swiss with high education.

The rows denoted rail, bus and car present average values of the variables for each travel means. From these values we learn that there is no "perfect" transportation means. Due to zero out-of-vehicle time the mean car journey (25.9 minutes) is more flexible and faster than the mean rail (42.5 minutes) or bus (36.8 minutes) journey, but costs around 4.7 to 6 times as much. Not surprisingly, commuters face a trade-off between time and money cost when choosing an alternative.

Furthermore, there is seemingly strong positive association between car and the characteristics age, male, married, and kids, while the association between car and part-time work is negative. One explanation might be that auto use is typically positively related to personal income[9], which is positively related to, for example, age yet negatively related to working part-time.

[9]Income is missing in the census.

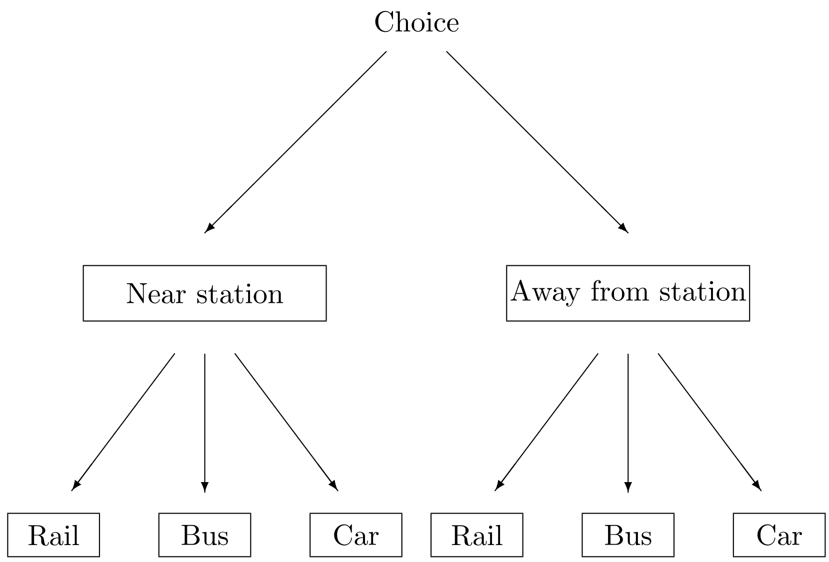


Figure 5.1: Joint Decision of travel mode and home location

5.7.3 Long-Term Sample

Theoretically, there is a limitless number of ways to define the choice of residence in terms of qualitative alternatives. In this work, the home location must fulfill the following criteria: First, it must be closely related to the choice of traffic mode. That is, the phenomenon of commuters' residential self-selection must be clearly recognizable in the data. Second, it must be binary so that the total number of choice alternatives is not too big and simulation-assisted estimation procedures can be applied. Third, it must be easily identifiable in the census data.

Following the idea of Cervero and Duncan (2002), we define the home location variable as follows: The commuter lives either *near the railway station* or *away from the railway station*. More precisely, near the railway station implies that the distance between residence and next station is less than or equal to 500 meters, while away from the railway station means that the distance is more than 500 meters. This distance makes sense from a content point of view, since it is covered easily and quickly by walking, on the bike or also by other transport means. In order to determine the home location variable from the data, we compare each person's home location coordinates in the census with the data of the Zurich Traffic Association containing the exact coordinates of all 175 train stations.

Altogether, the choice set comprises six alternatives of joint travel means and home location: {rail/rail, near rail/bus, near rail/car, away from rail/rail, away from rail/bus, and away from rail/car}. Figure

5.1 depicts the choice situation graphically. Note that the chronology of the decisions given by the decision tree is arbitrary and could be inverted (i.e. first the traffic mode and then the home location). For our research, the chronology known to us is not important since we regard the decisions as simultaneous (Koppelman, 1989).

5.7.4 Distribution of Long-Term Sample

Since the long-term sample comprises the same commuters as the short-term sample, the distribution of the sample observations across the travel alternatives rail, bus and car is as before: 50.02 percent rail, 7.80 percent bus, and 42.18 percent car. Table 5.4 displays that there are 30.73 percent observations residing near the railway station and 69.27 percent observation residing away from the railway station. As expected, the residential self-selection is clearly recognizable. Train riders rather live near the station in order to shorten the travel time and to reduce the travel costs, while car drivers and bus riders rather live away from the station. Therefore, we assume that the chosen residence choice is suitable for an investigation of the long-term traffic means demand.

In the residence definition, we have indicated the distance between home and station with 500 meters. Before, we argued that this distance is chosen wisely from a content point of view. It shows that it is also sensible from an empirical view. Other distances, such as for example 400 meters or 600 meters, the self-selection phenomenon is distinctly less pronounced for this sample.

The decisive question is if the residential self-selection is statistically significant. If not, we would be forced to define a new residence variable. We perform a homogeneity test in order to examine the null hypothesis that the travel mode distributions are independent of the home location. The chi^2-test statistic is 770.42 (with 2 degrees of freedom), so that we can reject the null on the one percent significance level.

In the table, the summary statistics are given for both residential alternatives. Not surprisingly, rail out-of-vehicle time scores are subject to the home location. On average, they are around 5 minutes shorter when the commuter lives near the railway station. This does not seem to be much. However, we have to consider that the out-of-vehicle time may be regarded as very unpleasant, for example in rainy or cold weather. Likewise, the average trip costs and the monthly rent are generally higher for people residing away from the station.

Furthermore, the characteristics age, male, part-time, male$\times$part-time, foreigner, low education, foreigner$\times$low education, medium education, foreigner$\times$medium education, high education, and

Table 5.4: Distribution of Long-Term Sample

	Near Rail				Away from Rail			
	Rail	Bus	Car	All	Rail	Bus	Car	All
Frequency	8,785	858	5,128	14,771	15,263	2,889	15,151	33,303
Sample Share (in %)	59.47	5.81	34.72	30.73	45.83	8.67	45.50	69.27
Out-of-vehicle time	16.06	11.14	0	9.07	21.83	10.10	0	10.64
	(4.30)	(7.52)	(0)	(8.38)	(5.26)	(5.36)	(0)	(9.92)
On-vehicle time	22.53	25.46	25.57	24.52	22.34	26.77	26.03	25.05
	(10.59)	(9.88)	(7.39)	(9.49)	(9.79)	(8.97)	(8.07)	(9.18)
Total cost	2.66	2.28	13.79	6.24	3.13	2.38	14.18	6.57
	(1.73)	(1.09)	(4.49)	(6.05)	(2.12)	(1.20)	(4.87)	(6.24)
Age	37.47	36.86	41.24	38.75	37.25	38.51	41.74	39.40
	(11.88)	(11.88)	(11.79)	(11.99)	(12.18)	(12.59)	(11.75)	(12.21)
Male	0.49	0.41	0.68	0.55	0.50	0.41	0.66	0.57
	(0.50)	(0.49)	(0.47)	(0.5)	(0.5)	(0.49)	(0.47)	(0.50)
Part-time	0.48	0.47	0.36	0.44	0.48	0.49	0.38	0.43
	(0.50)	(0.50)	(0.48)	(0.50)	(0.50)	(0.50)	(0.49)	(0.50)
Male×part-time	0.17	0.14	0.18	0.17	0.18	0.14	0.18	0.17
	(0.38)	(0.35)	(0.38)	(0.38)	(0.38)	(0.35)	(0.38)	(0.38)
Married	0.43	0.43	0.53	0.46	0.44	0.49	0.57	0.50
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Children	0.36	0.38	0.49	0.41	0.38	0.42	0.53	0.45
	(0.48)	(0.49)	(0.50)	(0.49)	(0.49)	(0.49)	(0.50)	(0.50)
Married×children	0.29	0.30	0.41	0.33	0.31	0.35	0.44	0.37
	(0.45)	(0.46)	(0.49)	(0.47)	(0.46)	(0.48)	(0.50)	(0.48)
Foreigner	0.17	0.29	0.21	0.19	0.16	0.26	0.19	0.18
	(0.38)	(0.45)	(0.41)	(0.39)	(0.37)	(0.44)	(0.39)	(0.38)
Low education	0.12	0.22	0.11	0.12	0.12	0.23	0.11	0.13
	(0.32)	(0.41)	(0.31)	(0.32)	(0.32)	(0.42)	(0.31)	(0.34)
Foreigner×low education	0.05	0.14	0.07	0.06	0.04	0.13	0.06	0.06
	(0.22)	(0.35)	(0.26)	(0.24)	(0.20)	(0.34)	(0.24)	(0.24)
Medium education	0.56	0.52	0.54	0.55	0.56	0.50	0.55	0.55
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Foreigner×medium education	0.07	0.09	0.09	0.08	0.07	0.07	0.08	0.07
	(0.26)	(0.29)	(0.29)	(0.27)	(0.26)	(0.26)	(0.27)	(0.26)
High education	0.33	0.26	0.35	0.33	0.32	0.28	0.34	0.32
	(0.47)	(0.44)	(0.48)	(0.47)	(0.47)	(0.45)	(0.47)	(0.47)
Foreigner×high education	0.06	0.06	0.05	0.06	0.05	0.05	0.06	0.06
	(0.24)	(0.24)	(0.22)	(0.24)	(0.22)	(0.22)	(0.24)	(0.24)
Monthly rent	1471	1409	1551	1495	1487	1459	1569	1522
	(487)	(524)	543)	(511)	(518)	(527)	(566)	(543)

Notes: Calculations are based 48,074 observations. Standard deviations in parentheses.

foreigner×high education do not vary strongly across the home location, suggesting that they will only have few influence on the residence choice. On the other hand, married, children, or the interaction married×children seem strongly associated with home the location away from the station.

Chapter 6

Short-Term Travel Mode Demand

6.1 Introduction

The aim of this chapter is to study commuters' short-term choices from among the travel alternatives rail, bus and car using MNL, ECMNL and RPMNL models. In these models, travel demand will be specified and examined at the individual level and then subjected to aggregation in order to predict and evaluate the behavior of the population in response to policy scenarios. The expression *short-term* implies that the commuter treats long-term decisions closely related to the mode choice as a fixed durable. The data sample used to realize this study has been summarized previously.

As discussed in Part I, the MNL model is the most basic model to quantify the relationship between a qualitative choice variable and a set of explanatory variables. It exhibits several advantages, including simplicity of estimation, yet its usefulness is limited because of the restrictive IID error term assumption that leads to the presence of the IIA property. On the contrary, ECMNL and RPMNL models, both belonging to the MMNL family, are most flexible to describe worker's travel decision realistically, yet require simulation assisted estimation procedures. The ECMNL model is typically used when we suspect heteroscedasticity and correlation in the unobserved portion of utilities. Moreover, by means of the RPMNL model we generalize the MNL by allowing for unobserved taste heterogeneity in the population.

6.2 Short Review of Previous Literature

The analysis of short-term travel mode demand has a long history in transportation economics and traffic planning, reflected by the mass of empirical studies written in past decades. Though many of these studies have contributed substantially to our understanding of the subject, there is still considerable work to be done in the future.

Up to the 1970s, travel forecasts were based on linear regression models, so-called gravity models, which describe aggregated route choice between origin and destination (Beckmann et al., 1956). However, unable to capture causal effects, zone-based models revealed only low degree of prediction accuracy. For further discussion of these early concepts, see Meyer and Straszheim (1971) and Boyce and Williams (2005).

Attempts to develop individual travel demand models started to appear in the mid 1970 with the derivation of the MNL and the NMNL model. Using a sample collected in the San Francisco Bay Area before the Bay Area Rapid Transit (BART) rail system was constructed in the mid 1970s, McFadden (1974, 1978) and Domencich and McFadden (1975) pioneered by employing MNL models to provide forecasts for travel alternatives including BART. Table 6.1, taken from McFadden's 1978 paper, illustrates the prediction success. The BART market share estimated by the MNL model was 6.4 percent, close to the actual share measured after the BART introduction in 1975, which was 6.2 percent. Presumably because of the IIA property, the MNL model slightly tended to overestimate public transport, while substantially underestimating the automobile portion. However, McFadden's figures were considerably more precise than the ones provided by aggregated zone-based regression models that prognosticated 15 percent BART share.

Later, comparisons confirmed the generally superior predictive power of disaggregated qualitative choice approaches of travel demand (see for example Watson and Westin, 1975). Since then, RUM based modeling has become the dominant paradigm to describe travel behavior (Ben-Akiva and Lerman, 1985). See McFadden (2001) for a 30-year retrospective review.

Table 6.1: Prediction Success Table for Pre-BART Model and Post-BART Data

	Car Alone	Bus/Walk	Bus/Car	BART/Bus	BART/Car	Car Pool
Predicted Share	55.8%	12.5%	2.4%	1.1%	5.3%	22.9%
(Standard Error)	(11.4%)	(3.4%)	(1.4%)	(0.5%)	(2.4%)	(10.7%)
Actual Share	59.9%	10.8%	1.4%	1.0%	5.2%	21.7%
Prediction Error	-4.1%	1.7%	1.0%	0.1%	0.1%	1.2%

Source: McFadden, 1978.

Though the IIA property has not always been properly justified, early research in transportation modeling saw an overwhelming use of the MNL and NMNL error structures due to the closed-form expressions available for calculating choice probabilities. These models were preferred to the more general but computationally less tractable approaches over a long time (see for example, Train, 1980; Hensher, 1986; Swait and Ben-Akiva, 1987; Bhat, 1997; de Palma and Rochat, 2000; Tiwari and Kawakami, 2001; Brownstone et al., 2003; Liu, 2006).

Nevertheless, as a result of improved computer capacity and speed, the use of analytically unsolvable integral-form models has become increasingly more popular. Especially the findings obtained from MMNL models emphasized the need to generalize the limited MNL and NMNL models in order to evaluate traffic policy strategies more realistically. Many studies using RPMNL models found evidence for substantial unobserved preference heterogeneity, suggesting that estimated probability elasticities are sensitive to violations of the fixed parameter assumption of the MNL (for example, Algers et al, 1998; Bhat, 1998a, 2000a; Alpizar and Carlsson, 2001; Hole and FitzRoy, 2004; Shen et al., 2005; Cirillo and Axhausen, 2006). Likewise, studies employing ECMNL approaches to examine the zero off-diagonal terms of the MNL variance-covariance matrix showed that non-IID error terms lead to considerable improvements in the model fit (Bhat, 1998b, 2000b).

Despite of a fair amount of research, there have been urgent identification and normalization issues associated with ECMNL models that have not been understood so far (Walker et al., 2004, 2006). As a result, researchers decided to impose a priori structures on the variance-covariance matrix instead of examining the full expression. To the author's knowledge, the study in this chapter is therefore the first providing the estimation outcome of the entirely identified and normalized ECMNL error portion.

6.3 Estimation of the Multinomial Logit Model

Initially we estimate and test MNL models of the commuter's modal decision for home-to-work-to-home trips. The ML estimation results of the final specification are summarized in Table 6.2, including indirect utility parameter estimates, the log-likelihood value at convergence, and the likelihood ratio index that indicates a very good model fit. The first column of the table states the variables of the observed utility portion, and the other three columns give the parameter estimates together with the standard errors (in parentheses).

In the paragraphs below, we discuss some of the salient findings. Primarily, the significance of the effects of single variables and groups of variables is discussed by variable categories. Based on sample size, we can expect ML estimation to follow its asymptotic properties and test statistics to follow their asymptotic distributions.

6.3.1 Level-of-Service Attributes

Consider the set of policy-sensitive level-of-service attributes, which are by far best investigated in literature. The recognition of their importance is apparent since the beginning of modern travel demand research.

First, we address the issue of finding the appropriate functional form to improve the model fit. In previous studies, indirect utilities were linear functions of time and cost measures usually. However, constant marginal utility is a restrictive assumption, which is probably inadequate. In order to obtain a somewhat richer specification, we develop indirect utility functions motivated by polynomials of third degree. The statistical results of this approach are highly encouraging. For the majority of attributes, the third-degree polynomial is statistically reliable on the one, five or ten percent significance level. Nevertheless, polynomials with one or more insignificant parameters need additional examinations. For rail out-of-vehicle time, bus out-of-vehicle time, and bus trip cost we compare the cubic with the linear and quadratic specification. As likelihood ratio test statistics show in Table 6.9 (see appendix), the cubic form is statistically superior on the one, five or ten percent significance level.

For interpretation, we must consider all three terms of the polynomial. To simplify matters, we calculate the ranges where the marginal utilities with respect to the attributes are negative. As Table 6.3 reveals, marginal utilities regarding to on-vehicle time are generally negative in short or in long journeys, while in between travelers obviously attach positive weight on each additive minute. In the latter case, one may argue that extra travel time can be used to keep working or to relieve tension (Redmond and Mokhtarian, 2002). More out-of-vehicle time, by contrast, is accompanied with growing time aversion. Thus, rail and bus riders usually perceive walking or waiting time as a burden. Moreover, the U-form regarding rail and car cost implies that commuters evaluate expenditures positively as long as cost are low or very high.

Finally, we turn to the question whether some or all mode attributes can be measured generically rather than alternative-specifically. Table 6.10 (see appendix) gives the corresponding log-likelihood values at convergence and likelihood ratio test statistics. In detail, we can reject two out of three

Table 6.2: MNL Model with Alternative-Specific Level-of-Service Parameters

	Rail	Bus	Car
Constant	-4.794***	-7.100***	$0^{\dagger}$
	(0.7858)	(0.7976)	($0^{\dagger}$)
On-vehicle time	-0.499***	-0.235***	-1.432***
	(0.0162)	(0.0183)	(0.0341)
Squared on-vehicle time	0.0149***	0.0047***	0.0427***
	(0.0007)	(0.0007)	(0.0011)
Cubed on-vehicle time	-0.00010***	-0.00002**	-0.00037***
	(0.00001)	(0.00001)	(0.00001)
Out-of-vehicle time	0.001	-0.002	$0^{\dagger}$
	(0.0297)	(0.0304)	($0^{\dagger}$)
Squared out-of-vehicle time	-0.0031**	-0.0040**	$0^{\dagger}$
	(0.0013)	(0.0018)	($0^{\dagger}$)
Cubed out-of-vehicle time	0.00003*	0.00005*	$0^{\dagger}$
	(0.00002)	(0.000028)	($0^{\dagger}$)
Total cost	0.762***	-0.422***	0.3317***
	(0.0420)	(0.1387)	(0.1135)
Squared total cost	-0.1033***	0.0343	-0.0153***
	(0.0059)	(0.0381)	(0.0055)
Cubed total cost	0.0037***	-0.0012	0.0002*
	(0.0002)	(0.0029)	(0.00009)
Age	-0.116***	-0.087***	$0^{\dagger}$
	(0.0071)	(0.0111)	($0^{\dagger}$)
Squared age	0.0011***	0.0009***	$0^{\dagger}$
	(0.0001)	(0.0001)	($0^{\dagger}$)
Male	-0.571***	-1.028***	$0^{\dagger}$
	(0.0347)	(0.0549)	($0^{\dagger}$)
Part-time	0.149***	0.341***	$0^{\dagger}$
	(0.0386)	(0.0557)	($0^{\dagger}$)
Male×part-time	0.095*	-0.024	$0^{\dagger}$
	(0.0504)	(0.0812)	($0^{\dagger}$)
Married	-0.004	-0.010	$0^{\dagger}$
	(0.0379)	(0.0612)	($0^{\dagger}$)
Children	-0.192***	-0.272***	$0^{\dagger}$
	(0.0485)	(0.0804)	($0^{\dagger}$)
Married×children	0.014	0.209**	$0^{\dagger}$
	(0.0595)	(0.0980)	($0^{\dagger}$)
Foreigner	-0.491***	0.302***	$0^{\dagger}$
	(0.0714)	(0.0906)	($0^{\dagger}$)
Medium education	-0.289***	-0.634***	$0^{\dagger}$
	(0.0508)	(0.0718)	($0^{\dagger}$)
Foreigner×medium education	0.259***	-0.317***	$0^{\dagger}$
	(0.0841)	(0.1158)	($0^{\dagger}$)
High education	-0.113**	-0.518***	$0^{\dagger}$
	(0.0537)	(0.0789)	($0^{\dagger}$)
Foreigner×high education	0.423***	0.071	$0^{\dagger}$
	(0.0894)	(0.1243)	($0^{\dagger}$)

Number of observations$= 48,074$
Log-likelihood$= -33,373.81$
Likelihood Ratio Index$= 0.37$

Notes: Standard errors in parentheses. † indicates a restriction to this fixed value. ***, **, * indicate levels of significance of one, five, and ten percent.

Table 6.3: Negative Marginal Utility

Attribute	Ranges
Rail on-vehicle time	$[0, 21.1)$, $(81.4, \infty)$
Bus on-vehicle time	$[0, 30.2)$, $(148.0, \infty)$
Car on-vehicle time	$[0, 24.6)$, $(53.0, \infty)$
Rail out-of-vehicle time	$(0.2, 65.8)$
Bus out-of-vehicle time	$(0, 51.4)$
Rail total cost	$(5.1, 13.5)$
Bus total cost	$[0, \infty)$
Car total cost	$(13.8, 50.6)$

null hypotheses, namely that travelers evaluate rail, bus and car on-vehicle time equally (one percent significance level) and that travelers attach the same weight on rail, bus and car cost (one percent significance level). However, the null hypothesis that out-of-vehicle time parameters do not vary across rail and bus riders cannot be rejected, indicating that aversion against waiting or walking is not contingent on the transportation means.

6.3.2 Probability Elasticities

In addition to the statistic significance, the question arises if the estimated effects of the level-of-service attributes are economically significant. To find this out, empirical research has focused on determining mean own point elasticities of the probabilities to quantify the shift of travel in an alternative that occurs due to policy strategies. Table 6.4 shows the predicted values calculated by using sample enumeration. In line with what was found for other countries, most short-term demand functions are rather inelastic (McFadden, 1974; Bhat, 2000a; Swait and Ben-Akiva, 1987; Alpizar and Carlsson, 2001; Hole and FitzRoy, 2004; Shen et al., 2005; Liu, 2006). Thus, commuters are expected to show little short-term responsiveness to changes in policy-sensitive attributes.

In detail, the table shows that rail riders are quite sensitive to changes in out-of-vehicle time. As a result, policy strategies with the objective to reduce the time required to and from the station, for example through improved bus shuttle services, would be quite successful in supporting commuters to shift their demand to rail. On the contrary, monetary and onboard time incentives having only a small or even negative impact may not be effective to boost rail ridership.

Furthermore, traffic policy in the canton of Zurich has a wide scope for possible actions to encourage more commuters to use the bus. These strategies especially include bus out-of-vehicle time reductions through a more frequent service and more extensive route coverage, bus onboard time reductions

Table 6.4: Mean Own Probability Elasticities

With respect to	Rail	Bus	Car
On-vehicle time	0.16	−0.21	0.80
Out-of-vehicle time	−0.94	−0.65	
Total cost	0.23	−0.57	−1.01

through the introduction of express services, and the subsidy of fares. Consequently, traffic policy aiming at increasing the efficiency of the transport system should invest in the bus net.

Of the incentives that could be used to discourage car ridership, an increase of travel cost is expected to be most effective[10]. A hypothetical increase in the car cost by 10 percent, for example induced by a rise in petrol taxes, would reduce expected car usage by more than 10 percent. This confirms our a priori expectation that increasing the cost of driving is an effective deterrent to car use, but contradicts previous research finding that car demand in Switzerland is much more elastic to changes in on-vehicle time than in travel cost (Vrtic et al., 2000). Furthermore, it is found that commuters show positive sensitivity to the length of time onboard, implying that they do not regard time delay caused by back-ups as a problem.

Why are some elasticities positive? Let us again look at Table 6.3. The functional form of the level-of-service attributes allows for the existence of positive areas of the demand function, namely wherever the marginal utility of an additional unit of the attribute is positive. Depending on the areas in which the most observations are located, the elasticity values in Table 6.4 are positive or negative. In addition, mean values are sensitive to outliers. If you calculated the medians instead of the means, you receive significantly lower values: −0.05 for rail on-vehicle time, 0.31 for rail cost, and 0.3 for car cost.

6.3.3 Socio-demographic Characteristics

In Table 6.2, the second set of explanatory variables comprises individual characteristics (including interactions of individual characteristics) and alternative-specific constants. Keep in mind that the rail and bus parameters are estimated and tested relative to the base category car whose parameters are fixed to zero.

[10]Traffic policy discussions always talk about strategies to reduce car traffic, however, never about strategies to increase it. Although we do not want to argue in favor of or against one traffic means, we agree with the prevailing argumentation at this point.

Table 6.5: Mean Probability Effects

	Rail	Bus	Car
Age	-0.017	-0.002	0.019
Squared age	0.012	0.002	-0.014
Male	-0.065	-0.050	0.114
Part-time	0.024	0.018	-0.032
Male×part-time	0.017	-0.005	-0.013
Married	0.001	-0.001	-0.000
Children	-0.024	-0.011	0.036
Married×children	-0.005	0.014	-0.009
Foreigner	-0.097	0.041	0.056
Medium education	-0.028	-0.032	0.060
Foreigner×medium education	0.055	-0.026	-0.029
High education	-0.003	-0.029	0.032
Foreigner×high education	0.071	-0.010	-0.061

The t-scores of the alternative-specific constants unveil high statistical significance (one percent significance level). This is an indication that some important predictors are missing in our analysis. Signs and values of the constants have no interpretation. Consistent with our expectations, most commuter characteristics and interaction terms have statistically high significant impacts on utility functions (one percent significance level), suggesting that there is indeed observed choice heterogeneity in the population. Presumably because of the small share of bus riders in the sample, the significance of the bus parameters is weaker than the one of rail and car parameters.

For interpretation, Table 6.5 gives average probability effects, determined by sample enumeration and ordered by characteristics. Overall, the values have the expected signs, but are rather small. This means that socio-economic control variables only have a slight influence on the choice of transport means. In other words, choice heterogeneity is statistically significant, however, economically not as relevant as level-of-service attributes[11].

On an average, age probability effects are U-shaped for public transportation means and inversely U-shaped for car. Thus, the young and the old are more likely to commute by rail and bus relative to car and to travelers in midlife. One possible explanation might be that our estimates capture effects of variables that are not considered in the model. For example, young people might favor public transport due to their stronger ecological awareness. Moreover, strong car demand in midlife might result from a growing income in this period of life.

[11]This statement is also supported by the fact that the likelihood ratio value of the model without socio-economic characteristics is 0.34. By contrast, the value without level-of-service attributes only is 0.21. Hence, travel mode choice mainly reacts to changes in level-of-service attributes.

The effect of a dummy variable usually depends on the effect of another dummy variable, and including an interaction term helps revealing the impacts of all possible dummy combinations. More concretely, we can estimate the single effects of each dummy as much as the effect of the presence of both dummies, always relative to the base category that represents the absence of both dummies.

First, consider the pair *male* and *part-time*. The mean probability effects of *male and full-time*, *female and part-time*, and *male and part-time* are calculated relatively to the base category *female and full-time*: -0.065, 0.024, and -0.024 for rail, -0.050, 0.012, and -0.037 for bus, and 0.114, -0.045, and 0.095 for car. These numbers indicate that males are more likely to drive but less likely to ride the train and the bus (relative to females), especially when they work full-time (relative to part-time). This interpretation seems credible because part-time jobs are normally associated with lower income, which can be negatively related to car availability (Pendyala et al., 1995).

As regards *marriage*, which is considered together with the presence of *children*, the average probability effects of *married without children*, *not married with children*, and *married with children*, relative to the reference group *not married without children*, are summarized as follows: 0.001, -0.024, and -0.028 for rail, -0.005, -0.011, and 0.002 for bus, and -0.000, 0.036, and 0.027 for car. It can be seen that the presence of children generally decreases the likelihood of rail and bus, yet increases the likelihood of car, whereas, by contrast, choice probabilities are nearly unaffected by marriage alone. Nonetheless, marriage lowers the children's impact on bus and car, while rising the children's impact on rail.

Finally, we explore the question whether the Swiss are actually more likely to travel by rail than foreigners. In this context, we are confronted with the potential association of people's nationality and education. According to the table, estimated mean probability effects of the categories *foreigner with low education*, *Swiss with medium education*, *foreigner with medium education*, *Swiss with high education*, and *foreigner with high education* relative to the base group *Swiss with low education* are as follows: -0.097, -0.028, -0.70, -0.003, and -0.023 for rail, 0.041, -0.032, -0.017, -0.029, and 0.002 for bus, and 0.056, 0.060, 0.087, 0.032, and 0.025 for car. Not surprisingly, medium and high education are positively associated with car use (relative to low education). This seems reasonable since we expect education to be positively related to income. Furthermore, foreigners are generally more committed to use the bus and the automobile, but not the rail (relative to the Swiss). However, choice heterogeneity caused by origin vanishes with more education, and high-qualified foreigners behave similar to the Swiss.

6.3.4 Hausman-McFadden Test

Though the present outcomes seem plausible, caution is necessary because the MNL model hypothesizes the presence of proportional substitution patterns between the alternatives known as IIA property. When this condition is violated, however, we risk discarding valuable information supplied by the data and drawing false conclusions from the commuter's choice behavior.

There are several possibilities to examine whether the MNL model is appropriate in this choice situation. A quite informal test is to peer whether the MNL probability estimates deviate from the relative frequencies in the sample. From Chapter 5 we know that the distribution of the sample observations across the three work-trip modes is 50.02 percent rail, 7.80 percent bus, and 42.18 percent car. On average, the values provided by the MNL model are 50.02, 7.79, and 42.19, respectively. At first glance, these figures suggest that the MNL model is suitable.

Let us take a closer look. We check the validity of the IIA property by carrying out the Hausman-McFadden test. Referring to equation (3.13), we obtain p-values of the test statistics that are close to one for both restricted choice sets $\tilde{C}_1$ ={rail,car} and $\tilde{C}_2$ ={bus,car}, indicating that the assumptions of the MNL model cannot be rejected by the data. These findings encourage our belief that the reported probability elasticities and probability effects are actually valid.

An important note must be made here. Some previous studies attempted to avoid the IIA issue by reducing the multinomial choice set to a binary set containing automobile and public transport only (for example, Vrtic et al., 2000; de Palma and Rochat, 2004). This approach causes the following problems: First, there is a loss of valuable information about the choice process when rail and bus do not share the same indirect utility parameters. For the current sample, we can reject the null hypothesis that the parameters of both alternatives are equal ($\chi^2 = 4416.78$, 38 degrees of freedom). Second, a common disturbing term for the public transport means rail and bus is only methodically correct if both disturbing terms correlate fully. However, in our case, the disturbing terms are independent and combining the alternatives would be methodically questionable.

6.4 Estimation of Mixed Multinomial Logit Models

The Hausman-McFadden test does not provide definitive statistical evidence whether the IIA property is supported by the data or not. Only when the non-IID random factors of more flexible models are insignificant, it is safe to say that the MNL predictions above are unbiased. We therefore examine

Table 6.6: ECMNL Model

$\hat{\omega}_{22}$	0.004
	(0.0242)
$\hat{\omega}_{23}$	0.007
	(0.0300)
$\hat{\omega}_{33}$	0.001
	(0.0545)
Number of observations$= 48,074$	
Log-likelihood$= -33,373.33$	
Likelihood Ratio Index$= 0.37$	

Notes: Standard errors in parentheses.

the indirect utility specification within the scope of ECMNL and RPMNL error frameworks, which are most appropriate to reduce the limitations of the MNL.

6.4.1 Error Component Multinomial Logit Model

The ECMNL model permits an increased level of flexibility in the specification of the variance-covariance matrix of utilities. For this reason, we must attach great importance on the correct identification and normalization of the model. Recall from (4.9) that for $J = 3$ outcomes the transformation variance-covariance matrix contains exactly three terms labeled ω_{22}, ω_{23}, and ω_{33} that can be estimated. In doing simulation assisted estimation procedures, fundamental aspects of the convergence analysis are the type of draws and the number of iterations of the simulation process. In accordance to Bhat (2003) and Train (1999), the MSL estimations are made using 100 Halton draws.

Table 6.6 gives the central estimation results, including the estimates of the ω-terms, the log-likelihood value at convergence, and the likelihood ratio index. We dispense with reporting the estimated indirect utility parameters because they are very similar to the ones of the MNL model. In line with the Hausman-McFadden test before, none of the ω-expressions differs statistically significantly from zero. We perform a likelihood ratio test to investigate the null hypothesis stating that the terms are jointly zero. The value of the test statistic is 0.96, indicating that the transformation variance-covariance matrix of the ECMNL model does not vary statistically significantly from that of the MNL.

6.4.2 Random Parameter Multinomial Logit Model

Continuing our attempt to find that error specification which provides the best model fit, now we make use of RPMNL models. As explained in Chapter 4.3.10, there may be unobserved taste variation in the population even if the IID error term assumption cannot be rejected. In such models, the specification of indirect utilities is the same as in the MNL, with the exception that the taste parameters on on-vehicle time, out-of-vehicle time, and trip cost plus the alternative-specific constants are allowed to vary normally distributed over observations rather than being fixed[12]. For analytical simplicity, we let the random parameters arise from an *independent* multivariate normal distribution.

The goal of estimating RPMNL models is to obtain information about the population distribution of each taste parameter. The vector of utility parameters is estimated using MSL with 100 Halton draws. In Table 6.7, we report the important results, including the mean and the standard deviations of the independent normal distribution, the log-likelihood value at convergence, and the likelihood ratio index. The parameter estimates of the commuter characteristics are very similar to those in Table 6.2 and are therefore not stated again.

The most striking result is that only two out of twenty-four unobserved standard deviations are statistically significantly different from zero (on the ten percent significance level). However, t-tests do not account for correlation between parameters, and we examine the whole RPMNL specification as well as reduced versions by employing a series of likelihood ratio tests. In Table 6.8, the statistics of these tests suggest that none of the RPMNL versions is statistically superior to the MNL. Thus, there is no evidence for unobserved taste variation in the population.

6.4.3 Remarks

It is not astonishing that the results obtained from MMNL models do not vary statistically significantly from the MNL outcome. Since we were not able to reject the null hypothesis of the Hausman-McFadden test, we did not expected to find patterns of unobserved correlation and unobserved taste variation in the data. The MMNL analyses have only further confirmed that the MNL assumptions are valid and that we accept the MNL forecasts as correct.

[12] The random specification of the constants is motivated by the belief that the constants capture the effects of missing level-of-service attributes.

Table 6.7: RPMNL Model

		Rail	Bus	Car
Constant	Mean	-4.972***	-7.100***	$0^{\dagger}$
		(0.7866)	(0.7984)	($0^{\dagger}$)
	Std. deviation	0.0121	0.0156	$0^{\dagger}$
		(0.0109)	(0.0178)	($0^{\dagger}$)
On-vehicle time	Mean	-0.499***	-0.234***	-1.433***
		(0.0162)	(0.0183)	(0.0341)
	Std. deviation	0.0006	0.0006	0.0004
		(0.0005)	(0.0006)	(0.0004)
Squared on-vehicle time	Mean	0.0149***	0.0047***	0.043***
		(0.0007)	(0.0007)	(0.0011)
	Std. deviation	0.0000	0.0000	0.0000
		(0.0000)	(0.0000)	(0.0000)
Cubed on-vehicle time	Mean	-0.00010***	-0.00002**	-0.00037***
		(0.0000)	(0.0000)	(0.0000)
	Std. deviation	0.0000	0.0000	0.0000
		(0.0000)	(0.0000)	(0.0000)
Out-of-vehicle time	Mean	0.001	-0.003	$0^{\dagger}$
		(0.0297)	(0.0304)	($0^{\dagger}$)
	Std. deviation	0.0004	0.0011	$0^{\dagger}$
		(0.0005)	(0.0017)	($0^{\dagger}$)
Squared out-of-vehicle time	Mean	-0.0031**	-0.0039**	$0^{\dagger}$
		(0.0013)	(0.0018)	($0^{\dagger}$)
	Std. deviation	0.0000	0.0000	$0^{\dagger}$
		(0.0000)	(0.0001)	($0^{\dagger}$)
Cubed out-of-vehicle time	Mean	0.00003*	0.00005*	$0^{\dagger}$
		(0.00002)	(0.0000)	($0^{\dagger}$)
	Std. deviation	0.0000	0.0000	$0^{\dagger}$
		(0.0000)	(0.0000)	($0^{\dagger}$)
Total cost	Mean	0.766***	-0.422***	0.333***
		(0.0420)	(0.1391)	(0.1136)
	Std. deviation	0.0000	0.0145*	0.0004
		(0.0022)	(0.0075)	(0.0006)
Squared total cost	Mean	-0.1040***	0.0343	-0.0153***
		(0.0059)	(0.0382)	(0.0055)
	Std. deviation	0.0002	0.0005	0.0000
		(0.0002)	(0.0020)	(0.0000)
Cubed total cost	Mean	0.0037***	-0.0012	0.0002*
		(0.0002)	(0.0029)	(0.0001)
	Std. deviation	0.0000*	0.0002	0.0000
		(0.0000)	(0.0003)	(0.0000)

Number of observations= $48,074$
Log-likelihood= $-33,362.15$
Likelihood Ratio Index= 0.37

Notes: Standard errors in parentheses. † indicates a restriction to this fixed value. ***, **, * indicate levels of significance of one, five, and ten percent.

Table 6.8: Likelihood Ratio Test Statistics

Random vector on	Log-Likelihood	Test Statistic	DF
All taste parameters	$-33,362.15$	23.32	26
Constants	$-33,372.83$	1.96	2
On-vehicle time parameters	$-33,368.31$	11.00	9
Out-of-vehicle time parameters	$-33,369.85$	7.92	6
Trip cost parameters	$-33,369.80$	8.02	9

6.5 Summary and Conclusions

This chapter has investigated commuters' short-term choices from among the travel alternatives rail, bus and car in the canton of Zurich. In doing so, it has treated all relevant decisions linked with the mode as exogenous. Overall, the results of this study may be important to decision-makers in traffic policy and planning wishing to predict commuters' short-term responsiveness to changes in traffic policy strategies.

In conclusion, we offer the following recommendations. In brief, the best way to reduce the number of cars on the streets is to increase car expenditures. Seen to be sensitive to out-of-vehicle time, short-term rail market share is increased effectively by improving shuttle services to railway stations. Moreover, short-term bus demand can be increased by more extensive bus route coverage and by introducing long-distance buses, which could be a fast and flexible competitor to rail and automobile.

This study aimed at contributing to the growing transportation literature by attaching importance on some relevant aspects arising in travel modeling. These included developing the final model through a systematic process of estimating and testing various indirect utility and error term specifications. As it turned out, the findings provided evidence for the current sample that the IID error term restriction imposed by the MNL model could not be rejected. As a result, the outcomes obtained from ECMNL and RPMNL models did not differ statistically significantly from that obtained from the MNL model.

6.6 Appendix

Table 6.9: Test of Polynomial Degree

Attribute	Function	Log-Likelihood	Test Statistic	DF
Rail out-of-vehicle time	Linear	$-33,382.32$	17.02^{***}	2
	Quadratic	$-33,375.26$	2.9^{*}	1
Bus out-of-vehicle time	Linear	$-33,377.73$	7.84^{**}	2
	Quadratic	$-33,375.31$	3.00^{*}	1
Bus trip cost	Linear	$-33,376.93$	6.24^{**}	2
	Quadratic	$-33,375.31$	3.00^{*}	1

Notes: ***, **, * indicate levels of significance of one, five, and ten percent.

Table 6.10: Test of Generic Specifications

Generic Specification	Log-Likelihood	Test Statistic	DF
On-vehicle time	$-35,079.42$	3,411.22***	6
Out-of-vehicle time	$-33,375.26$	2.9	3
Trip cost	$-33,745.35$	743.08***	6

Notes: *** indicates levels of significance of one percent.

Chapter 7

Long-Term Travel Mode Demand

7.1 Introduction

So far, much has been analyzed and written in transportation literature about the relationship between workers' travel mode choices and explanatory variables such as travel time and travel cost. However, there is an aspect that generally has been left out, but that plays an important role in a comprehensive behavioral demand theory. Studies as the one performed in Chapter 6 are deficient in the way that they do not confront the association of a range of choices (Waddell, 2001). In this chapter, we want to close this gap.

More concretely, this study is dealing with so-called residential self-selection, which is one of the most important aspects in the commuter's long-term travel decision process. Residential self-selection is defined as the people's tendency to look for home locations that allow them to use their preferred travel mode on the journey to work. In Chapter 5, we have defined the home location decision by the alternatives residing *near the railway station* and residing *away from the railway station*. In addition, the joint set of travel mode and home location decisions comprises the alternatives near rail/rail, near rail/bus, near rail/car, away from rail/rail, away from rail/bus, and away from rail/car.

Methodically, we study commuters' long-term travel demand using MNL, RPMNL and ECMNL models. Although no unobserved heteroscedasticity and correlation were found in the previous chapter, we again use the models of the MMNL family since the choice situation is not the same as before and therefore we cannot exclude that the unobserved parts of the utility functions are not IID anymore.

7.2 Short Review of Previous literature

Although the recognition that travel patterns and home location are interrelated is not new, this theme has not been widely adopted, and empirical transportation literatur covering this topic is extremely rare. Previous studies have mostly investigated either the mode decision or the home location decision, but not both together. A reason for this neglect might be the fact that only few traffic databases are able to identify the commuter's home residence selection.

After all, some previous studies found evidence for residential self-selection in Canada and in the USA. They showed that those who dislike driving mainly settle in urban areas because of the obvious association of easy access to public transport infrastracture and travel speed and travel cost (for example, Stringham, 1984; JHK and Associates, 1987, 1989; Cervero, 1993, 1994, 1996; Frank and Pivo, 1994; Gerston & Associates, 1995; Schimek, 1996; Kitamura et al., 1997). Due to this finding, a recent review of Boanet and Crane (2001) argued that travel patterns are at least partly a result of people's decisions where to live, and that this needs to be accounted for when studying work-trip behavior. Otherwise, the commonly observed relationship between travel behavior and a set of explanatory variables does not so much reflect direct causalities.

The article by Abraham and Hunt (1997) was one of the first that presented the results of models allowing the simultaneity of choices. The authors used NMNL models to describe the selection of residence, workplace and travel mode for multi-worker households in Canada to examine gender differences in travel mode and location decisions. Among others, they found that location choice is more important for female than for male travel demand. With the aim of receiving improved forecasts of land use and travel, Cervero and Duncan (2002) developed a NMNL model of travel and residential demand. Studying the year-2000 travel data from the San Francisco Bay Area, they discovered that residential self-selection accounts for approximately 40 percent of the rail commute decisions.

Many empirical questions remain unanswered, depite some recent work on this topic. In particular, none of the mentioned studies has researched both the short and long-term traffic means demand and compared the results. We therefore cannot answer the question regarding new discoveries when integrating the residential self-selection into traffic means models.

A further problem is the econometric methods that are usually used. So far, joint qualitative choice models have been specified as NMNL, which seems a reasonable approach to deal with the nested model structure such as in Figure 5.1 (Ben-Akiva and Bowman, 1998; Yao and Morikawa, 2005;

Rivera and Tiglao, 2005; Vega and Raynolds-Feighan, 2005). The NMNL model has some advantages, including simplicity of simulation-free ML estimation. Yet the variance-covariance structure of utilities is too limited in order to capture the individual decision process realistically.

7.3 Estimation of the Multinomial Logit Model

The effects of the exogenous variables on the endogenous variable are shown in Table 7.1, with the log-likelihood value at convergence and the customary goodness of fit measure. In the table, there are six columns displaying the estimates of the indirect utility parameters with standard errors (in parentheses). The predictor variables in the first column include personal characteristics and third-degree polynomials of level-of-service attributes, plus a third-degree polynomial of the household's monthly rent, which is hypothesized to influence the residence choice only. Begining with level-of-service attributes, we discuss the most important findings by variable categories.

7.3.1 Level-of-Service Attributes

In this model, level-of-service parameters are allowed to vary across travel modes, yet not across residential alternatives. Many but not all of them are statistically significant on the 1%, 5% or 10% significance level. We therefore investigate first which functional form achieves the best model fit. Likelihood ratio test statistics in Table 7.7 (see appendix) confirm that the mode-specific specification is statistically superior to the generic one on the one percent significance level. In addition, Table 7.8 (see appendix) unveils that the use of third-degree polynomials is statistically reliable for most attributes, with few exceptions only. Using likelihood ratio tests, we cannot reject the squared function regarding bus on-vehicle time and the linear function regarding bus trip cost.

From a statistical point of view, we can conclude that the results of the long-term and the short-term demand model are very similar. The lower likelihood ratio index value in Table 7.1 does not signify less significance of the long-term model. This goodness of fit measure cannot compare models with different endogenous variables.

Next, we concentrate on proofing the economic relevance of these results and identifying the differences compared to the results of the previous chapter. First, it is striking that the rail out-of-vehicle time and all cost parameter estimates in Table 7.1 substantially differ from those in Table 6.2. However, we cannot state if the observed differences are statistically significant or not. Instead, we

Table 7.1: MNL Model of Travel Mode and Residence Choice

	Near Rail			Away from Rail		
	Rail	Bus	Car	Rail	Bus	Car
Constant	-5.350***	-10.122***	-0.922***	-4.087***	-8.663***	0†
	(0.4799)	(0.6038)	(0.2502)	(0.4566)	(0.4956)	(0†)
On-vehicle time	-0.482***	-0.215***	-1.431***	-0.482***	-0.215***	-1.431***
	(0.0162)	(0.0183)	(0.0338)	(0.0162)	(0.0183)	(0.0338)
Squared on-vehicle time	0.0141***	0.0040***	0.0426***	0.0141***	0.0040***	0.0426***
	(0.0008)	(0.0007)	(0.0011)	(0.0008)	(0.0007)	(0.0011)
Cubed on-vehicle time	-0.00009***	-0.00001	-0.00037***	-0.00009***	-0.00001	-0.00037***
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Out-of-vehicle time	-0.407***	-0.010	0†	-0.407***	-0.010	0†
	(0.0279)	(0.0309)	(0†)	(0.0279)	(0.0309)	(0†)
Squared out-of-vehicle time	0.0162***	-0.0036**	0†	0.0162***	-0.0036**	0†
	(0.0013)	(0.0017)	(0†)	(0.0013)	(0.0017)	(0†)
Cubed out-of-vehicle time	-0.00022***	0.00005*	0†	-0.00022***	0.00005*	0†
	(0.00002)	(0.00003)	(0†)	(0.00002)	(0.00003)	(0†)
Trip cost	0.926***	-0.330**	0.159**	0.926***	-0.330**	0.159**
	(0.0584)	(0.1390)	(0.0644)	(0.0584)	(0.1390)	(0.0644)
Squared trip cost	-0.168***	-0.0009	-0.0113***	-0.168***	-0.0009	-0.0113***
	(0.0113)	(0.0379)	(0.0042)	(0.0113)	(0.0379)	(0.0042)
Cubed trip cost	0.0083***	0.0012	0.0001	0.0083***	0.0012	0.0001
	(0.0006)	(0.0029)	(0.00008)	(0.0006)	(0.0029)	(0.00008)
Rent	-0.031	-0.031	-0.031	-0.2997	-0.2997	-0.2997
	(1.0097)	(1.0097)	(1.0097)	(0.8854)	(0.8854)	(0.8854)
Squared rent	-1.946***	-1.946***	-1.946***	-1.609***	-1.609***	-1.609***
	(0.5633)	(0.5633)	(0.5633)	(0.4923)	(0.4923)	(0.4923)
Cubed rent	0.359***	0.359***	0.359***	0.295***	0.295***	0.295***
	(0.0997)	(0.0997)	(0.0997)	(0.0865)	(0.0865)	(0.0865)
Age	-0.095***	-0.065***	-0.004	-0.122***	-0.095***	0†
	(0.0091)	(0.0212)	(0.0102)	(0.0080)	(0.0123)	(0†)
Squared age	0.0009***	0.0005**	0.0001	0.0011***	0.0010***	0†
	(0.0001)	(0.0003)	(0.0001)	(0.0001)	(0.0001)	(0†)
Male	-0.543***	-1.048***	0.139***	-0.559***	-0.766***	0†
	(0.0443)	(0.1022)	(0.0512)	(0.0394)	(0.0622)	(0†)
Parttime	0.229***	0.230***	0.003	0.211***	0.408***	0†
	(0.0474)	(0.0984)	(0.0589)	(0.0429)	(0.0625)	(0†)
Male×parttime	-0.005	-0.0834	-0.059	0.046	-0.121	0†
	(0.0641)	(0.1522)	(0.0745)	(0.0569)	(0.0918)	(0†)
Married	-0.046	-0.136	-0.145***	-0.043	-0.026	0†
	(0.0481)	(0.1150)	(0.0554)	(0.0431)	(0.0690)	(0†)
Children	-0.234***	-0.229	-0.093	-0.216***	-0.311***	0†
	(0.0621)	(0.1519)	(0.0661)	(0.0551)	(0.0910)	(0†)
Married×children	0.005	0.178	0.099	0.101	0.254**	0†
	(0.0763)	(0.1869)	(0.0825)	(0.0675)	(0.1103)	(0†)
Foreigner	-0.303***	0.510***	0.184*	-0.490***	0.307***	0†
	(0.0916)	(0.1657)	(0.1005)	(0.0805)	(0.1012)	(0†)
Medium education	-0.159**	-0.513***	0.064	-0.287***	-0.650***	0†
	(0.0644)	(0.1372)	(0.0791)	(0.0565)	(0.0798)	(0†)
Foreigner×medium education	0.088	-0.311	-0.095	0.283***	-0.380***	0†
	(0.1083)	(0.2083)	(0.1163)	(0.095)	(0.1316)	(0†)
High education	0.034	-0.384**	0.155*	-0.091	-0.505***	0†
	(0.0685)	(0.1519)	(0.0822)	(0.0600)	(0.0878)	(0†)
Foreigner×hig education	0.340***	-0.270	0.239*	0.342***	0.086	0†
	(0.1140)	(0.2362)	(0.1257)	(0.1017)	(0.1388)	(0†)

Number of observations= $48,074$
Log-likelihood= $-62,919.59$
Likelihood Ratio Index= 0.27

Notes: Standard errors in parentheses. † indicates a restriction to this fixed value. ***, **, * indicate levels of significance of one, five, and ten percent.

compute the ranges where the marginal utilities with respect to the attributes are negative. As Table 7.2 shows, the values with respect to on-vehicle time, out-of-vehicle time and trip cost are similar but not equal to the ones reported in Table 6.3. For rail and car cost, for example, the negative ranges start with lower values than before. Hence, this table hardly reveals new discoveries and only the determination of mean probability elasticities will show if the long-term and short-term traffic means demand differ.

Table 7.2: Negative Marginal Utility

Attribute	Ranges
Rail on-vehicle time	$[0, 21.2)$, $(88.7, \infty)$
Bus on-vehicle time	$[0, 30.4)$, $(215.9, \infty)$
Car on-vehicle time	$[0, 24.6)$, $(53.0, \infty)$
Rail out-of-vehicle time	$(0, \infty)$
Bus out-of-vehicle time	$(0, 48.2)$
Rail total cost	$(3.8, 9.7)$
Bus total cost	$[0, 9.9)$
Car total cost	$(8.2, 51.0)$

7.3.2 Probability Elasticities

Table 7.3 displays mean own point probability elasticities, this time computed for both home residences. For comparative reasons, short-term elasticities based on the outcome in Table 6.2 are stated in parentheses.

Table 7.3: Short-Term and Long-Term Mean Own Probability Elasticities

	Near Rail			Away from Rail		
	Rail	Bus	Car	Rail	Bus	Car
On-vhicle time	1.00 (0.09)	−0.20 (−0.20)	0.85 (0.72)	0.51 (0.19)	−0.19 (−0.19)	0.81 (0.83)
Out-of-vehicle time	−0.82 (−0.56)	−0.66 (−0.70)		−0.62 (−1.11)	−0.63 (−0.63)	
Total cost	0.26 (0.24)	−0.68 (−0.57)	−1.17 (−1.01)	0.20 (0.23)	−0.66 (−0.57)	−1.17 (−1.05)

Notes: Short-term values in parentheses.

The following general conclusion can be drawn: The rail and car demand is rather more elastic in the long term than in the short term. However, this discovery cannot be made for the bus demand. This is not surprising since we have already shown in Chapter 5 that the phenomenon of the residential self-selection, as defined in this work, is not greatly pronounced. One could speculate that a new definition of the home location would be helpful for the long-term bus demand. On the other hand, the bus only plays a subordinated role compared to the train and car, and therefore, additional considerations are not necessary.

In more detail, abandoning the assumption that the home residence distribution is exogenous, we can say that long-term rail demand is more sensitive regarding on-vehicle time than short-term rail demand. The residence locations thereby play an important role. Commuters with a higher sensitivity as to wanting and being able to use their travel time productively prefer to live close to a station in the long term.

This latter observation is supported by the elasticity values regarding out-of-vehicle time. According to the table, rail demand is more sensitive to out-of-vehicle time for commuters living near the station than for commuters living away from the station. It must, however, be noted that the long and short-term values are mirror images. The long-term demand elasticities are short-term undervalued for commuters living near the station and overvalued for commuters living away from the station. From this observation, we conclude for example that the introduction of shuttle buses to reduce rail out-of-vehicle time for people who live away from the station does not have the distinct effect one expects from the results in the previous chapter. Persons with a high out-of-vehicle time sensitivity tend to move close to the station, and the measure therefore has less success than expected.

A similar picture results for the car demand. Car demand tends to be more elastic long-term compared to short-term regarding trip cost, yet independent from the home location. With respect to on-vehicle time, short-term and long-term elasticities seem to deviate especially for commuters living near the station.

7.3.3 Rent Prices

Furthermore, the results highlight the importance of rent prices (measured in 1000 Swiss francs) when choosing the home location. The central findings can be summarized as follows. Third-degree polynomials provide by far the best model fit. Table 7.8 (see appendix) shows that both the linear and quadratic function can be rejected on the one percent significance level.

As expected, a rise in rent prices in a certain area leads to a decreasing number of people likely to live there. Expressed in numbers, if the rental prices near the station increase by one percent, the number of persons near the station decreases by 4.2 percent (ceteris paribus). Vice versa, if the rental prices away from the station increase by one percent, the number of persons living away from the station decreases by 3.1 percent (ceteris paribus). However, it has to be considered that rental prices balance supply and demand for residence space and therefore, the ceteris-paribus-condition is hardly applicable in reality. Expressed differently, a rental price increase near the station - no matter what causes it - would increase demand and price for apartments away from the station, until an equilibrium is found.

7.3.4 Socio-economic Characteristics

Estimated relative to the base category *away from rail/car*, numerous socio-economic control variables in Table 7.1 are statistically significant on the 10%, 5% or 1% significance level. The comparatively many non-significant effects regarding the bus alternatives can be ascribed to the fact that the bus was only chosen by a minority of the sample observations. In addition, the comparatively many non-significant effects with respect to the alternative *near rail/car* can be attributed to the similarity of the base category. This points towards the fact that the researched control variables rather influence the choice of traffic mode than the choice of residence.

Table 7.4: Mean Probability Effects

	Near Rail			Away from Rail		
	Rail	Bus	Car	Rail	Bus	Car
Age	-0.002	0.000	0.005	-0.015	-0.001	0.013
Squared age	0.001	-0.000	-0.004	0.011	0.002	-0.010
Male	-0.019	-0.011	0.043	-0.047	-0.035	0.070
Part-time	0.010	0.002	-0.012	0.014	0.015	-0.029
Male×part-time	-0.003	0.002	-0.006	0.013	-0.007	0.001
Married	-0.000	-0.002	-0.011	0.000	-0.001	0.012
Children	-0.010	-0.001	0.003	-0.015	-0.009	0.032
Married×children	-0.011	0.002	0.004	0.010	0.011	-0.016
Foreigner	-0.012	0.013	0.033	-0.089	0.030	0.024
Medium education	0.008	-0.005	0.010	0.021	-0.030	-0.027
Foreigner×medium education	-0.006	-0.006	-0.015	0.062	-0.022	-0.013
High education	0.015	-0.005	0.020	-0.013	-0.024	0.007
Foreigner×high education	0.021	-0.006	-0.032	0.047	-0.004	-0.026

The probability effects reported in Table 7.4 are quite marginal. Hence, the observable choice heterogeneity between the different groups in the commuting population is statistically significant; however, the choice behavior is influenced essentially by the different level-of-service attributes and the rental prices[13].

If you add the values of the alternatives with the same traffic mode, you can observe that the values

[13] Without socio-economic characteristics, the likelihood ratio value of the model is 0.25, while the corresponding value of the model without level-of-service attributes and rental prices is only 0.18.

are almost identical to the ones in Table 6.5. Therefore, the conclusions are more or less the same as in the previous chapter, and we refrain from extensively discussing them.

7.4 Estimation of Mixed Multinomial Logit Models

Can we trust the findings presented in the preceding pages or should we question them? In a MNL, the estimations are made under the controversial IIA property. If the IIA is violated, the estimation results are biased and more flexible qualitative choice models should be applied. However, the use of the Hausman-McFadden test to examine this property is not feasible because some alternatives share common taste parameters. Even without this test, we can analyze the validity of the MNL assumptions as follows: If ECMNL and RPMNL models do not achieve a significantly better model fit than the MNL, we may assume that the MNL results are valid.

7.4.1 Error Component Multinomial Logit Model

Let us consider the ECMNL model whose error components allow for all patterns of unobserved heteroscedasticity and correlation that exist in the context of a multivariate normal distribution. Following the logic of illustration (4.9), at $J = 6$ choice options the transformation variance-covariance matrix has a total of 15 identified terms, as listed in Table 7.5. According to the outcome, which is obtained using MSL with 100 Halton draws, none of these terms is statistically significant. Furthermore, the value of the likelihood ratio test statistic is 10.8, indicating jointly insignificant non-IID error terms.

7.4.2 Random Parameter Multinomial Logit Model

As argued previously, it is however possible that there is unobserved taste variation in the population without the ECMNL error terms being significant (see Chapter 4.3.10). Therefore, we will finally examine a model with an RPMNL error structure.

In Table 7.6, only 3 out of 24 standard deviations are statistically significantly different from zero. Besides, the standard deviations are insignificant for all groups of attributes. This suggests that there is no statistical evidence for unobserved taste heterogeneity in the population.

Table 7.5: ECMNL Model

$\hat{\omega}_{22}$	0.010
	(0.0668)
$\hat{\omega}_{23}$	0.003
	(0.0355)
$\hat{\omega}_{24}$	0.005
	(0.0212)
$\hat{\omega}_{25}$	0.003
	(0.0192)
$\hat{\omega}_{26}$	0.007
	(0.0565)
$\hat{\omega}_{33}$	0.001
	(0.0092)
$\hat{\omega}_{34}$	0.006
	(0.0344)
$\hat{\omega}_{35}$	0.001
	(0.0079)
$\hat{\omega}_{46}$	0.004
	(0.0386)
$\hat{\omega}_{44}$	0.003
	(0.0211)
$\hat{\omega}_{45}$	0.004
	(0.0590)
$\hat{\omega}_{46}$	0.011
	(0.0655)
$\hat{\omega}_{55}$	0.003
	(0.0148)
$\hat{\omega}_{56}$	0.009
	(0.0472)
$\hat{\omega}_{66}$	0.003
	(0.0534)
Number of observations$= 48,074$	
Log-likelihood$= -62,914.19$	
Likelihood Ratio Index$= 0.37$	

Notes: Standard errors in parentheses.

Table 7.6: RPMNL Model

		Near Rail			Not Near Rail		
		Rail	Bus	Car	Rail	Bus	Car
Constant	Mean	-5.351***	-10.121***	-0.922***	-4.089***	-8.663***	0†
		(0.4804)	(0.6043)	(0.2505)	(0.4569)	(0.4961)	(0†)
	Std. dev.	0.0175	0.0277	0.0175	0.0044	0.0202	0†
		(0.0122)	(0.0346)	(0.0150)	(0.0104)	(0.0196)	(0†)
On-vehicle time	Mean	-0.480***	-0.214***	-1.432***	-0.480***	-0.214***	-1.432***
		(0.0162)	(0.0182)	(0.0339)	(0.0162)	(0.0182)	(0.0339)
	Std. dev.	0.0000	0.0007	0.0002	0.0000	0.0007	0.0002
		(0.0005)	(0.0006)	(0.0004)	(0.0005)	(0.0006)	(0.0004)
Squared on-vehicle time	Mean	0.0140***	0.0039***	0.0426***	0.0140***	0.0039***	0.0426***
		(0.0008)	(0.0007)	(0.0012)	(0.0008)	(0.0007)	(0.0012)
	Std. dev.	0.0000*	0.0000	0.0000	0.0000*	0.0000	0.0000
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Cubed on-vehicle time	Mean	-0.00009***	-0.00001	-0.00037***	-0.00009***	-0.00001	-0.00037***
		(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
	Std. dev.	0.0000**	0.0000	0.0000	0.0000**	0.0000	0.0000
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Out-of-vehicle time	Mean	-0.408***	-0.011	0†	-0.408***	-0.011	0†
		(0.0279)	(0.0306)	(0†)	(0.0279)	(0.0306)	(0†)
	Std. dev.	0.0002	0.0017	0†	0.0002	0.0017	0†
		(0.0005)	(0.0017)	(0†)	(0.0005)	(0.0017)	(0†)
Squared out-of-vehicle time	Mean	0.0163***	-0.0036**	0†	0.0163***	-0.0036**	0†
		(0.0013)	(0.0017)	(0†)	(0.0013)	(0.0017)	(0†)
	Std. dev.	0.0000	0.0000	0†	0.0000	0.0000	0†
		(0.0000)	(0.0000)	(0†)	(0.0000)	(0.0000)	(0†)
Cubed out-of-vehicle time	Mean	-0.00023***	0.00005*	0†	-0.00023***	0.00005*	0†
		(0.00002)	(0.00003)	(0†)	(0.00002)	(0.00003)	(0†)
	Std. dev.	0.0000	0.0000	0†	0.0000	0.0000	0†
		(0.0000)	(0.0000)	(0†)	(0.0000)	(0.0000)	(0†)
Trip cost	Mean	0.926***	-0.331**	0.158**	0.926***	-0.331**	0.158**
		(0.0581)	(0.1389)	(0.0639)	(0.0581)	(0.1389)	(0.0639)
	Std.dev.	0.0048*	0.0028	0.0008	0.0048*	0.0028	0.0008
		(0.0029)	(0.0074)	(0.0007)	(0.0029)	(0.0074)	(0.0007)
Squared trip cost	Mean	-0.168***	-0.0009	-0.0113***	-0.168***	-0.0009	-0.0113***
		(0.0112)	(0.0376)	(0.0040)	(0.0112)	(0.0376)	(0.0040)
	Std.dev.	0.0000	0.0019	0.0000	0.0000	0.0019	0.0000
		(0.0005)	(0.0019)	(0.0000)	(0.0005)	(0.0019)	(0.0000)
Cubed trip cost	Mean	0.0083***	0.0012	0.0001	0.0083***	0.0012	0.0001
		(0.0006)	(0.0029)	(0.00008)	(0.0006)	(0.0029)	(0.00008)
	Std.dev.	0.0000	0.0003	0.0000	0.0000	0.0003	0.0000
		(0.0001)	(0.0003)	(0.0000)	(0.0001)	(0.0003)	(0.0000)

Number of observations$= 48,074$
Log-likelihood$= -62,906.31$
Likelihood Ratio Index$= 0.27$

Notes: Standard errors in parentheses. † indicates a restriction to this fixed value. ***, **, * indicate levels of significance of one, five, and ten percent.

7.4.3 Remarks

Although the endogenous variable and therefore the decision situation are not the same as in the previous chapter, the statistic results regarding the unobserved factors are almost identical. The findings imply that the MNL assumption of IID error terms is supported by the data. This might be surprising, since one might expect unobserved correlation between the alternatives. For example, one could assume that common unobserved factors might appear between the same traffic means that influence the commuter's decision process. However, it appears that unobserved terms - however specified - hardly yield an explanation.

7.5 Conclusion

Even though it is important for efficient policy-making, modeling commuters' long-term travel mode choices has been a relatively undeveloped area in transportation literature. This chapter has contributed to the discussion by proposing a model that describes the commuter's joint decision between travel alternatives rail, bus or car on the one hand and between the residential alternatives living near the station and living away from the station on the other hand. In doing so, one broad purpose was to demonstrate that residential self-selection effects long-term travel mode choice processes.

Summing up the findings, the model results show the importance of allowing that travel and home location are interrelated. As expected, commuters' short-term and long-term reaction to policy strategies varies, albeit the differences are small. In general, rail and car demand is slightly more elastic in the long term than in the short term. In this particular case, this is valid for rail commuters who live near the station. These observations influence traffic policies; these in turn must consider that some intended traffic policy measures only take effect after several years.

In contrast to former literature that usually modeled long-term choices as MNL or NMNL, we also used flexible ECMNL and RPMNL models to allow for all patterns of unobserved correlation and taste variation in the population. Nevertheless, we could not find any statistically significant discrepancies from the IID assumption of the MNL.

7.6 Appendix

Table 7.7: Test of Generic Specifications

Generic Specification	Log-Likelihood	Test Statistic	DF
On-vehicle time	$-64,680.38$	3,521.58***	6
Out-of-vehicle time	$-62,966.92$	94.66***	3
Trip cost	$-63,239.10$	639.02***	6

Notes: *** indicates levels of significance of one percent.

Table 7.8: Test of Polynomial Degree

Attribute	Function	Log-Likelihood	Test Statistic	DF
Bu on-vehicle time	Quadratic	$-62,920.47$	1.76	2
Bus trip cost	Linear	$-62,921.30$	3.42	2
	Quadratic	$-62,919.67$	0.16	1
Rent	Quadratic	$-62,927.64$	16.12^{***}	2
	Linear	$-62,929.11$	19.06^{***}	4

Chapter 8

Conclusions

This thesis has investigated commuters' short-term and long-term choices from among the travel alternatives rail, bus and car. In the short-term analysis, we have treated all relevant decisions linked with the travel mode decision as exogenous (Chapter 6). In order to examine the long-term demand, the central idea was to let people self-select themselves geographically (Chapter 7). The empirical studies in both chapters have in common that they were undertaken to improve our understanding in what the short and long-term commuting behavior differs.

The results of this study may be highly relevant to decision-makers in traffic policy and planning. In general, contemporary Swiss traffic policy aims at altering commuters' travel mode behavior during the rush hour periods through strategies that include monetary disincentives for car use, such as parking pricing schemes and petrol taxes, and time and monetary incentives for transit use, such as frequent transit services, extensive transit route coverage, and the subsidy of fares. In our models, the implementation of these policies can be represented by changes in the policy-sensitive attributes out-of-vehicle time, on-vehicle time, and trip cost. The mean own probability elasticity values calculated in the empirical chapters might thus be the starting point for substantive implications. To a greater or lesser extend, the short-term results are compatible with what has been found in prior literature. From the long-term results we learn that residential self-selection matters for explaining long-term travel patterns. In particular, rail and car demand is slightly more elastic in the long term than in the short term.

Furthermore, the study attached importance on relevant aspects arising when modeling individual travel patterns. For example, to the author's knowledge, this is the first book presenting the outcome of completely identified and normalized ECMNL models. Nevertheless, the estimation results provided evidence that the IIA property imposed by the MNL model could not be rejected, implying that McFadden's well-known red-bus-blue-bus dilemma was not present in both data samples. As a result, probability elasticity values obtained from ECMNL and RPMNL models did not differ from the ones obtained from the MNL model.

Bibliography

[1] Abraham, J. and J. Hunt (1997), "Specification and estimation of nested logit model of home, workplaces, and commuter mode choices by multiple-worker households", *Transportation Research Record* 1606, 17-24.

[2] Algers, S., P. Bergström, M. Dahlberg, and J. Lindqvist Dillen (1998), "Mixed logit estimation of the value of travel time", Working Paper.

[3] Alonso, W. (1964), *Location and Land Use. Toward a General Theory of Land Rent*, Harvard University Press, Cambridge, MA.

[4] Alpizar, F. and F. Carlsson (2001), "Policy implications and analysis of determinants of travel mode choice: An application of choice experiments to Metropolitan Costa Rica", *Environment and Development Economics* 8. 603-619.

[5] Beckmann, M., C. McGuire, and C. Winston (1956), *Studies in the Economics of Transportation*, Yale University Press, New Haven.

[6] Ben-Akiva, M. and B. Francois (1983), "Mu-homogenous generalized extrem value model", Working Paper, Department of Civil Engineering, MIT.

[7] Ben-Akiva, M. and S. Lerman (1985), *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, MA.

[8] Ben-Akiva, M. and D. Bolduc (1996), "Multinomial probit with a logit kernel and a general parametric specification of the covariance structure", Working Paper, Department of Civil Engeneering, MIT.

[9] Ben-Akiva, M. and J. Bowman (1998), "Discrete choice methods and their applications in short term travel decisions", in R. Hall, ed., *Handbook of Transportation Science*, 2nd edition, Kluwer.

[10] Ben-Akiva, M. and M. Bierlaire (2003), "Discrete choice models with applications to departure time and route choice", in R. Hall, ed., *Handbook of Transportation Science*, 2nd edition, Kluwer.

[11] Bhat, C. (1997), Work travel mode choice and number of non-work commute stops", *Transportation Research B* 31, 41-54.

[12] Bhat, C. (1998a), "Accomodating variations in responsiveness to level-of-service measures in travel mode choice modeling", *Transportation Research A* 32, 495-507.

[13] Bhat, C. (1998b), "Accomodating flexible substitution patterns in multi-dimensional choice modeling: Formulation and application to travel mode and departure time choice", *Transportation Research B* 32, 455-466.

[14] Bhat, C. (2000a), "Incorporating observed and unobservec heterogeneity in urban work travel mode choice modeling", *Transportation Science B* 34, 228-238.

[15] Bhat, C. (200b), "A multi-level cross-classified model for discrete response variables", *Transportation Research B* 34, 567-582.

[16] Bhat, C. (2003), "Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences", *Transportation Research B* 37, 837-855.

[17] Bierlaire, M. (2001), "A theoretical analysis of the cross-nested logit model", Working Paper.

[18] Boarnet, M. and R. Crane (2001), "L.A. story: A reality check for transit-based housing", Working Paper, University of California.

[19] Borsch-Supan, A. and V. Hajivassiliou (1993), "Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models", *Journal of Econometrics* 58(3), pp. 347-68.

[20] Boyce, D. and C. Williams (2005), "Urban travel forecasting in the USA and UK", *Papers in Regional Science* 84, 85-103.

[21] Boyd, J. and J. Mellman (1980), "The effect of fuel economy standards on the US automotive market: A hedonic demand analysis", *Transportation Research A* 14, 367-378.

[22] Brownstone, D. and K. Train (1999), "Forecasting new product penetration with flexible substitution patterns", *Journal of Econometrics* 89, 109-129.

[23] Brownstone, D., D. Bunch, and K. Train (2000), "Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles", *Transportation Research B* 34, 315-338.

[24] Brownstone, D., A. Ghosh, T. Golob, C. Kazimi, and D. Van Amelsfort (2003), "Driver's willingness-to-pay to reduce travel time: Evidence from the San Diego I-15 congestion pricing project", *Transportation Research A* 37, 373-387.

[25] Bunch, D. (1991), "Estimability in the multinomial probit model", *Transportation Research B* 25, 1-12.

[26] Cardell, S. and F. Dunbar (1980), "Measuring the societal impacts of automobile downsizing", *Transportation Research A* 14, 423-434.

[27] Cervero, R. (1993), *Ridership impacts of transit-focused development in California*, Berkeley: Institute of Urban and Regional Development, University of California, Nonograph 45.

[28] Cervero, R. (1994), "Transit-based housing in California: Evidence on ridership impacts", *Transport Policy* 1(3), 174-183.

[29] Cervero, R. (1996), "Mixed land uses and commuting: Evidence from the American Housing Survey", *Transport Research A* 30, 361-377.

[30] Cervero, R. (2002), "Induced travel demand: Research design, empirical evidence, and normative policies", *Journal of Planning Literature* 17, 2-20.

[31] Cervero, R. and M. Duncan (2002), "Residential self selection and rail commuting: A nested logit analysis", Working Paper, University of California Transportation Center, Berkeley, CA.

[32] Chiou, L. and J. Walker (2006), "Masking identification of discrete choice models under simulation methods", Working Paper.

[33] Cirillo, C. and K. Axhausen (2006), "Evidence on the distribution of values of travel time savings from a six-week diary", *Transportation Research A* 40, 444-457.

[34] Daganzo, C. (1979), *Multinomial Probit: The Theory and Its Application to Demand Forecasting*, Academic Press, New York.

[35] De Palma, A. and D. Rochat (2000), "Mode choices for trips to work in Geneva: An empirical analysis", *Journal of Transport Geography* 8, 43-51.

[36] Domencich, T. and D. McFadden (1975), *Urban Travel Demand*, North Holland, Amsterdam.

[37] Eichenberger, R. and M. Schelker (2004), "Freie Fahrt, weniger Steuern", *Weltwoche*, Ausgabe 7/04.

[38] Frank, L. and G. Pivo (1994), "Impacts of mixed use and density on utilization of three modes of travel: Single-occupant vehicle, transit, and walking", *Transportation Research Record* 1466, 44-52.

[39] Gerston & Associates (1995), *Transit-Based Housing*, San Jose, Santa Clara County Transportation Agency and the Santa Clara Valley Manufacturing Association.

[40] Geweke, J. (1989), "Bayesian inference in econometric models using Monte Carlo integration", *Econometrica* 57(6), 1317-1339.

[41] Geweke, J., M. Keane, and D. Runkle (1994), "Alternative computational approaches to inference in the multinomial probit model", *Review of Economics and Statistics* 76, 609-632.

[42] Green, W. (2003), *Econometric Analysis*, 5th Edition, Prentice Hall.

[43] Hajivassiliou, V. (1990), "Smooth simulation estimation of panel data LDV models", Unpublished Manuscript, Yale University.

[44] Hajivassiliou, V. (1993), "Simulation estimation methods for limited dependent variable models", in G. Maddala, C. Rao, and H. Vinod, eds., *Handbook of Statistics*, Amsterdam and

New York, North-Holland, 1993, pp. 519-44.

[45] Hajivassiliou, V. and P. Ruud (1994), "Classical estimation methods for LDV models using simulation", in R. Engle and D. McFadden, eds., *Handbook of Econometrics*, Amsterdam and New York, North-Holland, 1994, pp. 2384-2441.

[46] Hajivassiliou, V., D. McFadden, and P. Ruud (1996), "Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results", *Journal of Econometrics*, 72(2), pp. 85-134.

[47] Halton, J. (1960), "On the efficiency of evaluating certain quasi-random sequences of points in evaluating multi-dimensional integrals", *Numerische Mathematika* 2, 84-90.

[48] Hammersley, J. and K. Morton (1956), "A new Monte Carlo technique: Antithetic variates", *Proceedings of the Cambridge Philosophical Society* 52, 449-474.

[49] Hausman, J. and D. Wise (1978), "A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences", *Econometrica* 48, 403-429.

[50] Hausman, J. and D. McFadden (1984), "Specification tests for the multinomial logit model", *Econometrica* 52, 1219-1240.

[51] Hensher, D. (1981), "Two contentions related to conceptual context in behavioral travel modelling", in P. Stopher, A. Meyburg, and W. Brög, eds., *New Horizons in Travel Behavior Research*, Lexington Books, Lexington, MA, pp. 509-514.

[52] Hensher, D. (1986), "Simultanous estimation of hierarchical logit mode choice models", Working Paper No. 24, MacQuarie University, School of Economic and Financial Studies.

[53] Hensher, D. and W. Green (2003), "The mixed logit model: The state of practice", *Transportation* 30, 133-176.

[54] Hess, S., M. Bierlaire, and J. Polak (2005), "Estimation of value of travel-time savings using mixed logit models", *Transporation Research A* 39, 221-236.

[55] Hess, S., J. Polak, and K. Train (2006), "The use of modified latin hybercube sampling (MLHS) method in the estimation of a mixed logit model of vehicle choice", *Transportation Research B* 40, 147-167.

[56] Hole, A. and F. FitzRoy (2004), "Commuting in small towns in rural areas: The case of St Andrews", Working Paper.

[57] Horowitz, J. and N. Savin (2001), "Binary response models: Logits, probits and semiparametrics", *Journal of Economic Perspectives* 15, 43-56.

[58] JHK & Associates (1987), *Development-Related Ridership Survey I*, Washington, D.C.: Washington Metropolitan Transit Authority.

[59] JHK & Associates (1989), *Development-Related Ridership Survey II*, Washington, D.C.: Washington Metropolitan Transit Authority.

[60] Keane, M. (1994), "A computationally practical simulation estimator for panel data", *Econometrica* 62(1), pp. 95-116.

[61] Kitamura, R., P. Mokhtarian, and L. Laidet (1997), "A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area", *Transportation* 24(2), 125-158.

[62] Koppelman, F. and C.-H. Wen (2000), "The paired combinatorial logit model: Properties, estimation and application", *Transportation Research B* 34, 75-89.

[63] Liu, G. (2006), "A behavioral model of work-trip mode choice in Shanghai", Discussion Papers No. 444, Statistics Norway, Research Department.

[64] Luce, D. (1959), *Individual Choice Behavior*, John Wiley and Sons, New York.

[65] Marschak, J. (1960), "Binary choice constraints on random utility indications", in K. Arrow, ed., *Stanford Symposium on Methematical Methods in the Social Sciences*, Stanford University Press, Stanford, CA, pp. 312-329.

[66] Manski, C. (1977), "The structure of random utility", *Theory and Decision* 8, 229-254.

[67] Manski, C. and S. Lerman (1981), "On the use of simulated frequencies to approximate choice probabilities", in C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, MIT Press, 305-19.

[68] McFadden, D. (1973), "Conditional logit analysis of qualitative choice behavior", in P. Zarembka, ed., *Frontiers in Econometrics*, Academic Press, New York, 105-142.

[69] McFadden, D. (1974), "The measurement of urban travel demand", *Journal of Public Economics* 3, 303-328.

[70] McFadden, D. (1978), "The theory and practice of disaggregate forecasting for various modes of urban transportation", in T. Oum, et al., eds., *Emerging Transportation Planning Methods*, U.S. Department of Transportation DOT-RSPA-50-78-2, Reprinted, Transportation Economics Selected Readings, Seoul, Seoul Press, 51-80.

[71] McFadden, D. (1989), "A methode of simulated moments for estimation of discrete response models without numerical integration", *Econometrica* 76(4), pp. 995-1026.

[72] McFadden, D. and K. Train (2000), "Mixed MNL models for discrete response", *Journal of Applied Econometrics* 15, 447-470.

[73] McFadden, D. (2001), "Disaggregate behavioral travel demand's RUM side: A 30-year retrospective", in D. Hensher, ed., *Travel Behavioral Research: The Leading Edge*, Pergamon, Amsterdam, 75-96.

[74] McGrath, E. (1970), *Fundamentals of Operations Research*, West Coast University Press, San Francisco.

[75] Meyer, J. and M. Straszheim (1971), "Techniques of transportation planning", The Brookings Institution, Washington.

[76] Moser, P. (2005), "Pendeln im Zürcher Wirtschaftsraum - ein Überblick", *Statistik Info* 02/2005, Statistisches Amt des Kantons Zürich.

[77] Pendyala, R., L. Kostyniuk, and K. Goulias (1995), "A repeated cross-sectional evaluation of car ownership", *Transportation* 22, 165-184.

[78] Revelt, D. and K. Train (1998), "Mixed logit with repeated choices: Households' choices of appliance efficiency level", *Review of Economics and Statistics* 80, 647-657.

[79] Rivera, M. and N. Tiglao (2005), "Modeling residential location choice, workplace location choice and mode choice of two-worker households in Metro Manila", *Proceedings of the Eastern Asia Society for Transportation Studies* 5, 1167-1178.

[80] Romani, J., J. Suriñach, and M. Artis (2003), "Are commuting and residential mobility decisions simultaneous? The case of Catalonia (Spain)", Working Paper.

[81] Schimek, P. (1996), "Household motor vehicle ownership and use: How much does residential density matter?", *Transportation Research Record* 775, 48-55.

[82] Schwanen, T. and P. Mokhtarian (2005), "What if you live in the wrong neighborhood? The impact of residential neighborhood type dissonance on distance traveled", *Transportation Research D* 10, 127-151.

[83] Schweizerischer Bundesrat (2002), "Strategie Nachhaltige Entwicklung", März 2002.

[84] Shen, J., Y. Sakata, and Y. Hashimoto (2005), "A comparison between latent class model and mixed logit model for transport mode choice: Evidences from two datasets of Japan", Discussion Paper.

[85] Stern, S. (1997), "Simulation based estimation", *Journal of Economic Literature* 35(4), 2006-39.

[86] Stringham, M. (1984), "Travel behavior associated with land uses adjacent to transit stations", *ITE Journal* 54(2), 18-22.

[87] Swait, J. and M. Ben-Akiva (1987), "Empirical test of a constraint choice discrete model: Mode choice in Brasil", *Transportation Research B* 21, 103-115.

[88] Thurstone, L. (1927), "A law of comparative judgement", *Psychological Review* 34, 273-286.

[89] Tiwari, P. and T. Kawakami (2001), "Modes of commuting in Mumbai: A discrete choice analysis", *Review Of Urban and Regional Development Studies* 13, 34-45.

[90] Train, K. (1980), "A structured logit model of auto ownership and mode choice", *Review of Economic Studies* XLVII, 357-370.

[91] Train, K. (1995), "Simulation methods for probit and related models based on convenient error partitioning", Working Paper No. 95-237, Department of Economics, University of California, Berkley.

[92] Train, K. (1998), "Recreation demand models with taste differences over people", *Land Economics* 74, 230-239.

[93] Train, K. (1999), "Halton sequences for mixed logit", Working Paper No. E00-278, Department of Economics, University of California, Berkeley.

[94] Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press.

[95] Train, K. and Z. Sandor (2004), "Quasi-random simulation of discrete choice models", *Transportation Research B* 38, 313-327.

[96] Tuffin, B. (1996), "On the use of low-discrepancy sequences in Monte Carlo methods", *Monte Carlo Methods and Applications* 2, 295-320.

[97] Vega, A. and A. Reynolds-Feigham (2005), "Choice of residential location and travel mode to work in the Greater Dublin Area", Working Paper.

[98] Vovsha (1997), "Application of a cross-nested logit model to mode choice in Tel Aviv, Israel, Metropolitan Area", *Transportation Research Record* 1607, 133-142.

[99] Vrtic, M., K. Axhausen, R. Maggi, and F. Rossera (2003), "Verifizierung von Prognosemethoden im Personenverkehr", im Auftrag der SBB und dem Bundesamt für Raumentwicklung (ARE), IVT, ETH Zürich und USI Lugano, Zürich und Lugano.

[100] Waddell, P. (2001), "Towards a behavioral integration of land use and transportation modelling", Paper Presented at the 9th International Association for Travel Behavior Research Conference, Queensland, Australia.

[101] Walker J. (2002), "The mixed logit (or logit kernel) model: Dispelling misconceptions of identification", Working paper.

[102] Walker, J., M. Ben-Akiva, and D. Bolduc (2004), "Identification of the logit kernel (or mixed logit) model", Working paper.

[103] Walker, J., M. Ben-Akiva, and D. Bolduc (2006), "Identification of parameters in normal error components logit mixture models", Working paper, Center for Transportation Studies, Boston University.

[104] Watson, P. and R. Westin (1975), "Transferability of disaggregate mode choice models", *Regional Science and Urban Economics* 5, 227-249.

[105] Williams, H. (1977), "On the formation of travel demand models and economic evaluation measures of user benefits", *Environment and Planning A* 9, 285-344.

[106] Yao, E. and T. Morikawa (2005), "A study of an integrated intercity travel demand model", *Transportation Research A* 39, 367-381.

Curriculum Vitae

Personal Details

Date and place of birth: 10th of June 1973, Entlebuch, Switzerland
Citizenship: Swiss

Education

2003–2007	Doctoral Studies in Economics at the University of Zurich, Supervisor: Prof. Dr. R. Winkelmann
2003/2004	Study Center Gerzensee: Program for Beginning Doctoral Students in Econometrics
2003	Licentiate in Economics, University of Zurich

Professional Experience

2003–2006	Research Associate, Statistics and Empirical Economics Research Group, Socio-economic Institute, University of Zurich
2005	Short-term consultancy at Swissfirst Bank: Development of Swiss Entrepreneurial Index (SENTX)

Publications

12/04	*Zeit und Geld. Wie beeinflussen Zeit und Kosten die Verkehrsmittelwahl von Zupendelden nach Zürich? In: Analysen 15/2004, Statistik Stadt Zürich.*